STATIC HAND GESTURE SEGMENTATION FOR IMAGES WITH COMPLEX BACKGROUND; DETECTION AND TRACKING OF DYNAMIC HAND GESTURE

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

MASTER OF TECHNOLOGY

IN

COMMUNCATION AND SIGNAL PROCESSING

BY

AVINASH BABU.D

211EC4091

UNDER THE GUIDANCE OF

PROF. SAMIT ARI

Department of Electronics and Communication Engineering
National Institute of Technology Rourkela-769008

2013

Department of Electronics & Communication Engineering

National Institute of Technology Rourkela

CERTIFICATE

This is to certify that the thesis entitled, "Color hand gesture segmentation for images with complex background" submitted by Mr. **D.AVINASH BABU** in partial fulfillment of the requirements for the award of Master of Technology Degree in Electronics and Communication Engineering with specialization in "**Communication and Signal Processing**" during session 2012-2013 at the National Institute of Technology, Rourkela is an authentic work carried out by him under my supervision and guidance. To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other University/ Institute for the award of any degree or diploma**.**

Prof. Samit Ari

Asst. Professor
Dept. of Electronics and Communication Engineering
National Institute of Technology Rourkela
Rourkela 769008

ACKNOWLEDGEMENTS

This project is by far the most significant accomplishment in my life and it would be impossible without people who supported me and believed in me.

I would like to extend my gratitude and my sincere thanks to my honorable, esteemed supervisor Prof. **Samit Ari**. He is not only a great teacher/professor with deep vision but also and most importantly a kind person. I sincerely thank for his exemplary guidance and encouragement. His trust and support inspired me in the most important moments of making right decisions and I am glad to work with him. My special thank goes to Prof**. S. Meher** Head of the Department of Electronics and Communication Engineering, NIT, Rourkela, for providing us with best facilities in the Department and his timely suggestions.

I want to thank all my teachers Prof**. S. K. Patra, K. K. Mahapatra, A. K. Sahoo, P. Singh and S. K. Behera** for providing a solid background for my studies and research. Thereafter, I would fail in my duty if I don't think the lab assistant **Mr. Kishore Kujur** without whom the work would not have progressed. They have been great sources of inspiration to me and I thank them from the bottom of my heart.

I would like to thank all my friends and especially my classmates for all the thoughtful and mind stimulating discussions we had, which prompted us to think beyond the obvious. I have enjoyed their companionship so much during my stay at NIT, Rourkela. I would like to thank all those who made my stay in Rourkela an unforgettable and rewarding experience.

Last but not least I would like to thank my parents, who taught me the value of hard work by their own example. They rendered me enormous support during the whole tenure of my stay in NIT Rourkela.

D. AVINASH BABU

CONTENTS

ABSTRACT

This thesis presents color hand gesture segmentation for static images with complex background along with tracking and detection of hand gesture from video sequence. This thesis consists of two works: 1) Static. 2) Dynamic.

In the first part, aim is to automatically segment the hand gesture from a given image under different luminance conditions and complex backgrounds. The luminance value affects the color component of an image which leads to increase the noise level in the segmented image. This paper proposes a combined model of two color spaces i.e., HSI, YCbCr and morphological operations with labeling to improve the segmentation performance of color hand gesture from complex backgrounds in terms of completeness and correctness. The proposed color model separates the chrominance and luminance components of the image. The performance of the proposed method is demonstrated through simulation and the experimental results reveal that proposed method provides better performance accuracy compared to the HSI and YCbCr methods individually in terms of correctness and completeness.

In the second part, aim is to automatic detection and tracking of hand gesture from video sequence under different backgrounds. It involves three steps: 1). Hand tracking 2). Hand detection 3). Hand identification.

Generally, all tracking system begins with motion detection. Motion detection separates the corresponding moving objects region from the background image. The first and foremost process in the motion detection is to capture the image information using a video camera. The motion detection stage involves basic image preprocessing step such as; gray-scaling and smoothing, reducing image resolution using low resolution image technique, frame difference, and morphological operations with labeling. The preprocessing steps are applied to decrease the image noise in order to achieve a better accuracy of the tracking.

Here all the simulations are done in MATLAB 10 environment.

LIST OF FIGURES

LIST OF TABLES

CHAPTER-1: INTRODUCTION

1.1 WHAT IS A DIGITAL IMAGE?

An image may be defined as a two dimensional function, f(x, y), where x and y are spatial coordinates, and the amplitude of ‘f ’at any pair of coordinates (x, y) is called the intensity or gray level of the image at that point. When x and y are the amplitude values of are all finite, discrete quantities, we call the image a digital image.

The field of digital image processing refers to processing digital images by means of a digital computer. Note that a digital image is composed of a finite number of elements, each of which has a particular location and value. These elements are referred to as picture elements, image elements, and pixels. Pixel is the term most widely used to denote the elements of a digital image.

There are no clear cut boundaries in the continuum from image processing at one end to computer vision at the other. However, one useful paradigm is to consider three types of computerized processes in this continuum: low-level, mid-level and high-level processes. Low level processing involves primitive operations such as image pre-processing to reduce noise, contrast enhancement and image sharpening. A low level processing is characterized by the fact that both its inputs and outputs are images. Mid- level processing on images involves tasks such as segmentation, description of those objects and classification of those objects. A midlevel processing is characterized by the fact that its inputs are images, but outputs are attributes extracted from the images. Finally high-level processing involves recognizing the objects as in image analysis.

1.2 DIGITAL IMAGE PROCESSING:

Image processing (IP) is a form of information processing.

I/P is an image-------→IP--------→ O/P not necessarily image

(photographs/frames of video) can be set of features of image

- A digital image is built up with a number of elements (pixels).

- The pixel values corresponds to its brightness or color in the picture.
- Brightness can be used to interpret different pictures in different ways.
- Brightness reflects the received signal strength
- Using Image processing target aspects are extracted.

Modern digital technology has made it possible to manipulate multi-dimensional signals with systems that range from simple digital circuits to advanced parallel computers. The goal of this manipulation can be divided into three categories:

- Image Processing image in -> image out
- Image Analysis image in -> measurements out
- Image Understanding image in -> high-level description out

We will focus on the fundamental concepts of image processing. Further, we will restrict ourselves to two-dimensional (2D) image processing although most of the concepts and techniques that are to be described can be extended easily to three or more dimensions.

The amplitudes of a given image will almost always be either real numbers or integer numbers. The latter is usually a result of a quantization process that converts a continuous range (say, between 0 and 100%) to a discrete number of levels. In certain image-forming processes, however, the signal may involve photon counting which implies that the amplitude would be inherently quantized. In other image forming procedures, such as magnetic resonance imaging, the direct physical measurement yields a complex number in the form of a real magnitude and a real phase.

1.3 IMAGE REPRESENTATION AND TYPES OF NOISE:

Interest in digital image processing methods stems forms two principle application Areas: improvement of pictorial information for human interpretation and processing of image data for storage, transmission, and representation for autonomous machine perception.

This chapter has several objectives: (1) to define the scope of the filed which we call image processing, (2) to define key terms like digital image, pixel and its representation, (3) type of images, (4) noise in digital images and its properties.

1.4 TYPES OF IMAGES:

Each pixel of an image is typically associated to a specific 'position' in some 2D region, and has a Value consisting of one or more quantities (samples) related to the position. Digital images can be classified according to the number and nature of those samples:

a) Binary
b) Gray Scale
c) Color
d) False-color

The term digital image is also applied to data associated to points scattered over a three dimensional region, such as produced by topographic equipment. In that case, each datum is called a vowel.

A) BINARY IMAGE

A binary image is a digital image that has only two possible values for each pixel. Binary images are also called bi-level or two-level. (the names black – and white, B&W, monochrome or monochromatic are often used for this concept, but may also designate any images that have only one sample per pixel, such as gray scale image.) Binary images often arise in digital image processing as masks or as the result of certain operations such as segmentation, thresholding, and dithering. Displays, can only handle bi-level images.

A binary image is usually stored in memory as a bitmap, a packed array of bits. Binary images can be interpreted as subsets of the two-dimensional integer lattice Z^2; the field of morphological image processing was largely inspired by this view.

B) GRAYSCALE

In computing, a gray scale or gray scale digital image is an image in which the value of each pixel is a single sample. Displayed images of this sort are typically composed of shades gray, varying from black at the weakest intensity to white at the strongest, though in principle the samples could be displayed as shades of any color, or even coded with various colors for different intensities. Gray scale images are distinct from black-and-white images, which in the context of computer imaging are images with only two colors, black and white; gray scale images have many shades of gray in between. In most contexts other than digital imaging, however, the term" black and white" is used in place of "gray scale"; for example, photography in shades of gray is typically called "black-and-white photography". The term monochromatic in some digital imaging contexts is synonymous with gray scale, and in some contexts synonymous with black-and-white.

Grayscale images are often the result of measuring the intensity of light at each pixel in a single band of the electromagnetic spectrum (e.g. visible light).Grayscale images are intended for visual display are typically stored with 8 bits per samples pixel, which allows 256 intensities (i.e., shades of gray) to be recorded, typically on a non-linear scale. The accuracy provided by this format is barely sufficient to avoid visible banding artifacts, but very convenient for programming. Technical uses (e.g. in medical imaging or remote sensing applications) often require more levels, to make full use of the sensor accuracy (typically 10 or 12 bits per sample) and to guard against round off errors in computations. Sixteen buts per sample (65536 levels) appear to be a popular choice for such uses. ..

C) COLOR IMAGE:

In this category, the images in question typically are acquired with a full color sensor, such as a color TV Camera or color scanner. The full color image processing techniques are now used in a broad range of applications, including publishing visualization and the internet.

D) FALSE – COLOR:

Pseudo color (also called false color) image processing consists of assigning colors to gray values based on a specified criterion. The term pseudo or false color is used to differentiate the process of assigning colors to monochrome images from the processes associated with true color images. The principle use of Pseudo color is for human visualization and interpretation of gray-scale events in an image or sequence of images.

The techniques adopted for gray scale image enhancement can be directly applicable to color images.

1.5 IMAGE REPRESENTATION:

An image is stored as a matrix using standard Mat lab matrix conventions. There are five basic types of images supported by Mat lab:

1. Intensity images
2. Binary images
3. RGB images
4. 8-bit images

These image types are primarily for the purpose of display. They do not constrain the values of an image that can be processes using general image processing techniques in Mat lab. An exception is uint8, the data type for 8-bit images. Ordinarily, image coordinates use the same conventions as matrix coordinates, with the first argument referring to row# and the second to column #. The origin is the upper left corner of the image. In Mat lab, the origin is (1, 1) when Mat lab asks for x and y coordinates, x is considered to be to the right and y is considered to be down

1.5.1 INTENSITY IMAGES:

An intensity image is a data matrix whose values have been scaled to represent intensities. When the elements of an intensity image are of class uint8, or class uint16, they have integer values in the range [0.255] and [0, 65535], respectively.

1.5.2 BINARY IMAGES:

Binary images have a very specific meaning in MATLAB. A binary image is logical array of 0s and 1s. Thus, an array of 0s and 1s whose values are of data class, say uint8, in not considered a binary image in MATLAB. A numeric array is converted to binary using function logical.

To test if an array is logical we use the islogical function:

Is logical (x)-if x is logical, this function returns 1.Otherwise it returns 0.

1.6 NOISE:

Real world signals usually contain departures from the ideal signal that would be produced by our model of the signal production process. Such departures are referred to as noise. Noise arises as a result of un-modeled processes going on in the production and capture of the real signal. It is not part of the ideal signal and may be caused by a wide range of sources, e.g. variations in the detector sensitivity, environmental variations, the discrete nature of radiation, transmission or quantization errors, etc. it is also possible to treat irrelevant scene details as if they are image noise (e.g. surface reflectance textures). The characteristics of noise depend on its source, as does the operator which best reduces its effects.

Many image processing packages contain operator to artificially add noise to an image. Deliberately corrupting an image with noise allows us to test the resistance of an image processing operator to noise and assess the performance of various noise filters.

1.7 TYPES OF NOISE:

Noise can generally be grouped into two classes:

- Independent noise.
- Noise which is dependent on the image data.

Image independent noise can often be described by an additive noise model, where the recorded image f (i, j) is the sum of the true images s (i, j) and the noise n (i, j)

$$F(i,j) = s(i,j) + n(i,j) \quad \ldots\ldots\ldots \quad (1.1)$$

The noise n (i, j) is often Zero-mean and described by its variance σ^2_n. The impact of the noise on the image is often described by the signal to noise ratio (SNR), which is given by

$$SNR = \frac{\sigma_s}{\sigma_n} = \sqrt{\frac{\sigma_f^2}{\sigma_n^2}} - 1 \quad \ldots\ldots\ldots\ldots \quad (1.2)$$

Where σ_s^2 and σ_n^2 are the variances of the true image and the recorded image, respectively.

In many cases, additive noise is evenly noise is evenly distributed over the frequency domain (i.e. white dominate for high frequencies and its mostly low frequency information. Hence, the noise is dominant for high frequencies and its effects can be reduced using some kind of low pass filter. This can be done either with a frequency filter or with a spatial filter. (Often a spatial filter is preferable, as it is computationally less expensive than a frequency filter.)

In this second case of data-dependent noise (e.g. arising when monochromatic radiation is scattered from a surface whose roughness is of the order of a wavelength, causing wave interference which results in image speckle), it is possible to model noise with a multiplicative, or non-linear, model. These models are mathematically more complicated; hence, if possible, the noise is assumed to be data independent.

1.7.1 DETECTOR NOISE:

One kind of noise which occurs in all recorded images to a certain extent is detector noise. This kind of noise is due to the discrete nature of radiation, i.e. the fact that each imaging system is recording an image by counting photons. Allowing some assumptions (which are valid for many applications) this noise can modeled with an independent, additive model, where the noise n (i, j) has a Zero-mean Gaussian distribution described by its standard deviation (**σ),** or

variance. This means that each pixel in the noisy image is the sum of the true pixel value and a random, Gaussian distributed noise value.

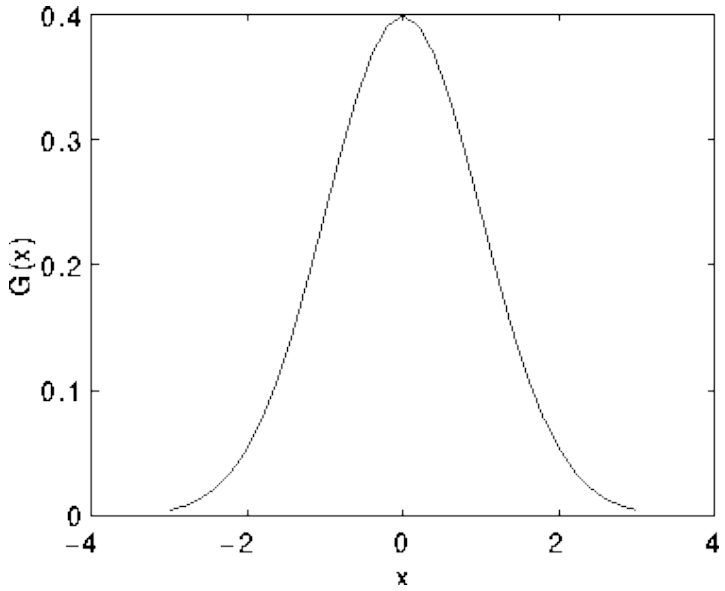


Figure1.1: Gaussian distribution with mean 0 and standard deviation 1.

1.7.2 SALT AND PEPPER NOISE:

Another common form of noise is data drop-out noise (commonly referred to as intensity spikes, speckle or salt and pepper noise). Here, the noise is caused by errors in the data transmission. The corrupted pixels are either set to the maximum value (which looks like snow in the image) or have single bits flipped over. In some cases, single pixels are set alternatively to zero or to the maximum value, giving the image a `salt and pepper' like appearance. Unaffected pixels always remain unchanged. The noise is usually quantified by the percentage of pixels which are corrupted.

1.8 EFFECT OF NOISE IN DIGITAL IMAGES:

In this section we will show some examples of images corrupted with Gaussian noise and give a short overview of which noise reduction operators are most appropriate.

1.8.1 GAUSSIAN NOISE:

We will begin by considering additive noise with a Gaussian distribution. If we add Gaussian noise with σ values of 8, we obtain the image.

Gaussian noise can be reduced using a spatial filter. However, it must be kept in mind that when smoothing an image, we reduce not only the noise, but also the fine-scaled image details because they also correspond to blocked high frequencies. The most effective basic spatial filtering techniques for noise removal include: mean filtering, median filtering and Gaussian smoothing. More sophisticated algorithms which utilize statistical properties of the image and/or noise fields exist for noise removal. For example, adaptive smoothing algorithms may be defined which adjust the filter response according to local variations in the statistical properties of the data.

1.9 CHARACTERISTICS OF NOISE:

- Images are often degraded by random noise. Noise can occur during image capture, transmission or processing, and may be dependent on or independent of image content.
- Noise is usually described by its probabilistic characteristics.

    - **White Noise** – constant its probabilistic characteristics.
      With increasing frequency); very crude approximation of image noise

    - **Gaussian noise** –It is a very approximation of noise that occurs in may practical cases
    - Probability density of the random variable is given by the Gaussian curve;
    - **ID Gaussian Noise** - μ is the mean and σ is the standard deviation of the random variable.

$$P(x) = \frac{1}{\sigma\sqrt{2\prod}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \ldots\ldots\ldots\ldots\ldots (1.3)$$

During image transmission, noise which is usually independent of the image signal occurs.

- Noise may be **Additive**, noise and image signal are independent.

$$f(x,y) = g(x,y) + v(x,y) \text{ ……..... (1.4)}$$

Multiplicative, noise is a function of signal magnitude.

$$f = g + vg = g(1+v) \approx gv \text{ …… (1.5)}$$

Impulse noise (saturated = salt and pepper noise).

1.10 INTRODUCTION TO DENOISING:

In many cases, additive noise is evenly distributed over the frequency domain (i.e., white noise), whereas an image contains mostly low frequency information. Hence, the noise is dominant for high frequencies and its effects can be reduced using some kind of low-pass filter. This can be done either with a frequency filter or with a spatial filter. Often a spatial filter is preferable, as it is computationally less expensive than a frequency filter.)

De-noising can be done in various domains and by using various methods

a) spatial domain
b) frequency domain

1.11 Spatial Domain Filters:

Spatial filters work directly on the pixels of an image. We use the term spatial filtering to differentiate this type of processing from the more traditional frequency domain filtering.

Traditional filters for image De-noising

a) Mean filters
b) Median filters, and
c) Gaussian smoothing filters

COMMON VALUES:

There are standard values for the various parameters encountered in digital image processing. These values can be caused by video standards, by algorithmic requirements, or by the desire to keep digital circuitry simple. Table 1 gives some commonly encountered values.

TABLE I : COMMON VALUES OF DIGITAL IMAGE PARAMETERS.

Parameter	Symbol	Typical values
Rows	N	256,512,525,625,1024,1035
Columns	M	256,512,768,1024,1320
Gray Levels	L	2,64,256,1024,4096,16384

The number of distinct gray levels is usually a power of 2, that is, $L=2^B$ where B is the number of bits in the binary representation of the brightness levels. When B>1 we speak of a gray-level image; when B=1 we speak of a binary image. In a binary image there are just two gray levels which can be referred to, for example, as "black" and "white" or "0" and "1".

1.12 THESIS MOTIVATION:

Hand gesture recognition is an essential task of today's research due to the increasing demands for human–computer interactions (HCIs) in recent years. Image segmentation is the process of partitioning an image into its constituent components i.e. homogeneous and meaningful regions acceding to their identical set of properties or attributes. Recently, there has been a high level of interest in recognizing human hand gestures. Hand-gesture recognition has so many applications like computer games, machinery control and thorough mouse replacement. One of the most ordered sets of gestures belongs to sign language. In sign language, each hand gesture is having an assigned meaning. Hand gestures can be categorized in two groups: static and dynamic. A static gesture is a particular hand configuration and pose, represented by a single image. A

dynamic gesture is a moving gesture, represented by a sequence of framing mages from a video. We will focus on the recognition of both static and dynamic images.

1.13 LITERATURE REVIEW:

Most of the researchers have proposed many methods for hand gesture recognition systems. Generally, such recognition systems are divided into two basic approaches namely:

1) Glove based approach.2) Vision based approach.

In glove based approach, detection of the hand is eliminated by the sensors on the hand. Such types of systems are minimal in use for body motion capture purposes and widely used in industry. On the other hand vision-based approach is more natural and used in real time applications. A human can easily identify a hand gesture, however for a computer to recognize and segment hand gesture first the hand should be detected in the acquired image and recognition of that hand should be done in a same way as humans do. Still this method is more challenging approach to execute because of the limitations of such a natural system. The vision-based approaches are generally carried out by using one or more than one camera to capture and analyze respective gestures of hands.

Detection and gesture analysis of the hands is an emerging literature topic and has many user environment limitations for most of the studies both in static and dynamic case. Segmentation of the hand gesture is the first step of such type of systems. Such type of systems should be able to identify the difference between complex backgrounds and hand images. So that it can segment the hand gesture region automatically and eliminate the unwanted background i.e. noise can easily eliminated. In recent past years, with the introduction of a new technique which has a high recognition rate, case studies are mostly concentrated on Boosting and HMM [1].But the drawback of this method is that it need straining and testing process. This process sometime needs a big-sized sample images to have a high detection and recognition rate. One more disadvantage is training process takes high computational time and it might be several days to complete training.

Many studies in literature use different skin thresholds for different color spaces to locate faces or hands in images by using different histogram methods [2]. In the RGB color space which is

primary color space the brightness and the color component of the images are not separated so Douglas Chai and King N. Ngan [3] proposed color segmentation using YCbCr color space. By combining the HSI and the YCbCr color space can easily separates the luminance and chrominance components of the image so this method can be applicable for different lightning condition and for different skin color under complex backgrounds [4]. Finally the performance of the proposed hybrid algorithm is evaluated in terms of completeness and correctness.

1.14 SCOPE OF THIS PROJECT:

The scope of this work is to automatically segments the hand gesture from a given image under different luminance conditions and complex backgrounds. The luminance value affects the color component of an image which leads to increase the noise level in the segmented image. Special care has to be taken at the time of segmentation to nullify the noise components in the image i.e. unwanted background of the image by applying morphological operations. This work presents a combined model of two color spaces i.e., HSI, YCbCr [5] and morphological operations with labeling to improve the segmentation performance of color hand gesture from complex backgrounds in terms of completeness and correctness.

The proposed color model separates the chrominance and luminance components of the image. The performance of the proposed method is demonstrated through simulation and the experimental results reveal that proposed method provides better performance accuracy compared to the HSI and YCbCr methods individually in terms of correctness and completeness.

1.15 THESIS OUTLINE:

The outline of this thesis is as follows.

Chapter 1: It presents the basic theory of Digital images and representation of images and noise level in images, types of noise, characteristics of noise.

Chapter 2: This chapter focuses on the study of skin color modeling analysis and different color spaces for modeling.

Chapter 3: This chapter describes the Color Hand gesture segmentation for images with complex background and explains the steps of proposed methodology.

Chapter 4: In this chapter, it explains about the Hand detection, tracking and identification from video sequence using Block matching algorithm.

Chapter 5: In this chapter, conclusion of the thesis is given in this chapter and the future work of the Hand segmentation is explained.

CHAPTER-2:

SKIN COLOR MODEL AND COLOR SPACES

2.1 INTRODUCTION TO SEGMENTATION METHODS:

Hand gesture recognition is an essential task of today's research due to the increasing demands for human–computer interactions (HCIs) in recent years. Image segmentation is the process of partitioning an image into its constituent components i.e. homogeneous and meaningful regions acceding to their identical set of properties or attributes. Segmentation algorithms are based on different parameters of an image like gray-level, color, texture, depth or motion [6]. The image segmentation process can be considered as one of the basic and important steps in digital image processing and computer vision applications such as tracking, pattern recognition and object identification. It is easy to distinguish the objects from the simple back ground but extraction of objects from the complex background of a digital image has been a challenging task in the field of digital image processing. With the increasing demand for complex image analysis and interpretation, the demand for accurate segmentation of images has also grown stronger and as a result many image segmentation methods and algorithms have been developed over the past few decades. The most popular method to perform image segmentation is gray level segmentation method [7] which is based on thresholding because it is simple and having a high speed of operation and ease of implementation.

However, the disadvantage of thresholding method is performance limited and suitable for only simple background images. All the color spaces are mathematical representation of a set of colors. All the color spaces are derived from the RGB information supplied by devices such as cameras and scanners. The most common are YCbCr, HSV, HSI, color spaces. The HSV (hue, saturation, value) color space is developed to be more intuitive in manipulating color and designed to approximate the way humans perceive and interpret color. The HSV color space is preferred for manipulation of hue and saturation i.e. to shift color or adjust the amount of color since it yields a greater dynamic range of saturation.

The HSI (hue, saturation and intensity) is similar to HSV model [8]. The main difference between these two models is the computing of the brightness component (I and V), which determines the distribution and dynamic range of both the brightness and saturation. The HSI method is best color space for the traditional image processing function like Convolution,

Equalization, and Histogram [9]. The YCbCr is another color space unlike the RGB color space, here the luminance or brightness or intensity is separated from the chrominance or pure color value [10]. The value of Y represents the luminance value and Cb and Cr represents the color or chrominance value, these are also known as color difference of the image [11].

Previous studies clearly intimate that people with different skin colors can be modeled by a skin models in different color spaces [12]. This can be done by taking different skin color values, from different users under various lightning conditions and a meaningful distribution is done to get closer to skin color pixels.

There are studies in literature to extract significant skin color boundaries to segment skin pixels in a given image effectively. These boundaries were designed to include all possible skin color values and named as skin color model. For HSV color space, a pixel is classified as skin color if the following conditions are satisfied otherwise non skin color.

$$0 < H < 52 \ldots\ldots\ldots\ldots \quad (2.1)$$

$$0.21 < S < 0.70 \ldots\ldots\ldots\ldots \quad (2.2)$$

For YCbCr color space, a pixel is classified as skin color if the following conditions are satisfied otherwise non skin color.

$$79 < Y \ldots\ldots\ldots\ldots\ldots \quad (2.3)$$

$$83 < Cb < 135 \ldots\ldots\ldots\ldots \quad (2.4)$$

$$132 < Cr < 180 \ldots\ldots\ldots\ldots \quad (2.5)$$

For Normalized RGB color space, a pixel is classified as skin color if the following conditions are satisfied otherwise non skin color.

$$g \leq r \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.6)$$

$$g \geq r - 0.4 \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.7)$$

$$g \geq -r + 0.6 \ldots\ldots\ldots\ldots\ldots\ldots \quad (2.8)$$

$$g \leq 0.4 \ldots\ldots\ldots\ldots\ldots \quad (2.9)$$

From the above equations, it is clear that the threshold values for the experiments are fixed. Each equation represents a border line for the skin color in the corresponding color space.

2.2 IMPORTANCE OF SKIN COLOR MODELING:

Skin color segmentation is an importance task in hand gesture system. The aim of skin color distribution is to find whether a pixel is skin color or non-skin color. Previous studies have implies that human skin color is not dependent on human race and the wavelength of the exposed light [13]. This observation clearly informs the necessity that the color space should be able to eliminate the luminance feature in proper way. Pure color information must be obtained and it must be not dependent on the brightness of the scene.

2.3COLOR SPACES:

There are different color spaces which are used for skin color modeling which derived from primary color space. All the color spaces are derived from RGB color space which is mathematical representation of set of colors. Different color spaces are: 1) RGB color space. 2) Normalized RGB color space. 3) YCbCr color space. 4) HSI color space.

2.3.1 RGBCOLOR SPACE:

The cone-sensors in human eye are mainly divided into three regions, namely Red, Green and Blue. RGB color space is primary color space and is mostly used in computer graphics. The main advantage of this color space is no conversion is required. In recent years, although other color spaces are well defined to such systems, RGB color space is still the most common color space to represent images. Most widely used method to have a robust system for changes in intensity is to trans-form RGB color space into another color space where chrominance and luminance components are orthogonal [14]. But the pure color information namely chrominance values of skin color are different in different color spaces and they might vary. The major disadvantage of

this method is this color space cannot separate chrominance and luminance components of the image. For this purpose, we study the other color spaces like YCbCr, HSI color spaces as these models can easily separates the luminance and chrominance components of the image from the color component of the image.

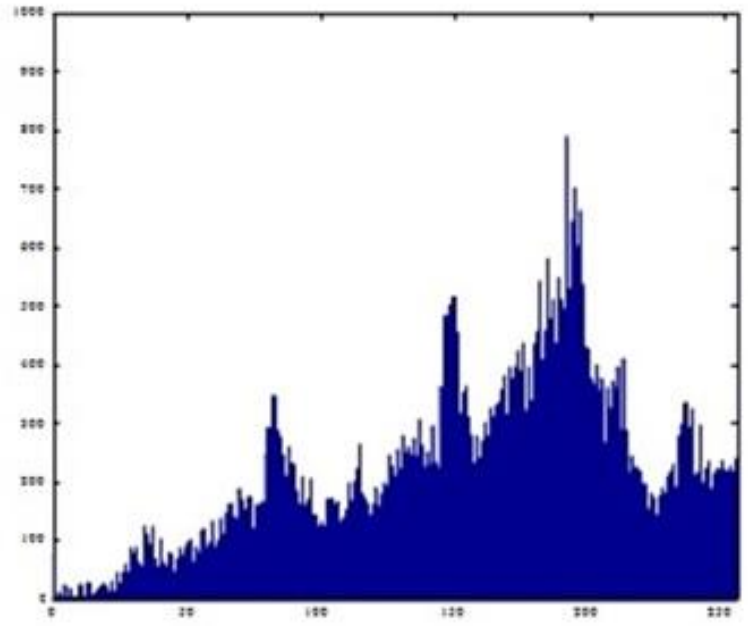


Figure 2.1: Histogram of R component for the skin color.

All the images in the computer are stored in RGB format. This is the greatest advantage of this color space. There is no conversion required for this segmentation method.

The main advantage of this method is it used the primary color space which decreases the computational complexity of the segmentation and the threshold value is dynamically calculated for each input. But the disadvantage is the RGB color space can't separate the luminance and the chrominance values.

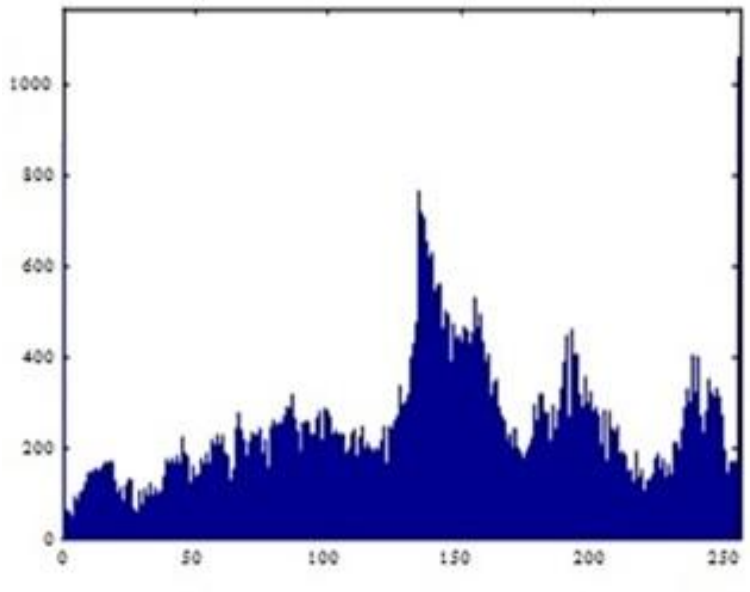


Figure 2.2: Histogram of G component for the skin color.

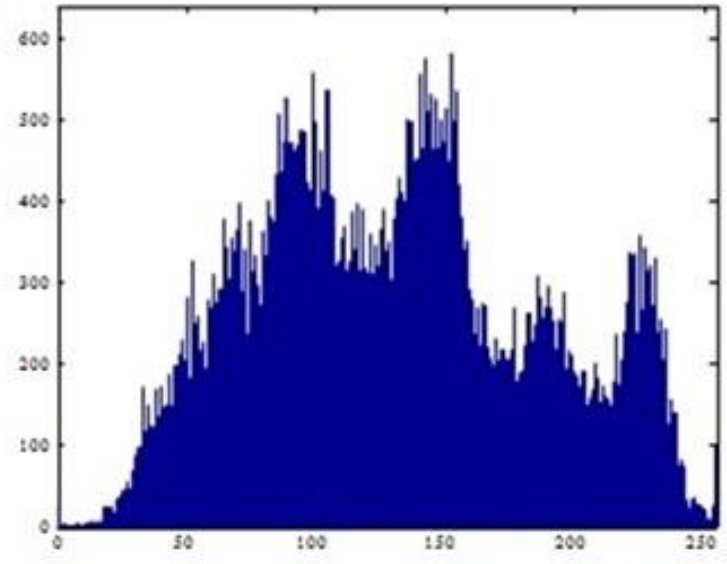


Figure 2.3: Histogram of B component for the skin color.

From the above figures, it is clear that every skin pixel is evenly and consistently distributed in RGB color space.

2.3.2 Normalized RGBCOLOR SPACE:

We know that RGB color space cannot separate the luminance and chrominance components from the color components of the image, unlike RGB color space Normalized RGB color space will separates the luminance and chrominance components which are orthogonal but not fully solve the problem. In this color space, all the values of RGB components will be normalized.

$$r = R / (R + G + B) \ldots\ldots\ldots\ldots \quad (2.10)$$

$$g = G / (R + G + B) \ldots\ldots\ldots\ldots \quad (2.11)$$

$$b = B / (R + G + B) \ldots\ldots\ldots\ldots \quad (2.12)$$

Normalized RGB color space also uses the same segmentation process as RGB color space. In Normalized RGB color space the luminance and chrominance effect is not totally eliminated but nullified [15].

This is a fast segmentation method for simple background. The complexity of this method is very low but this method can't be applicable for the bright images without any preprocessing.

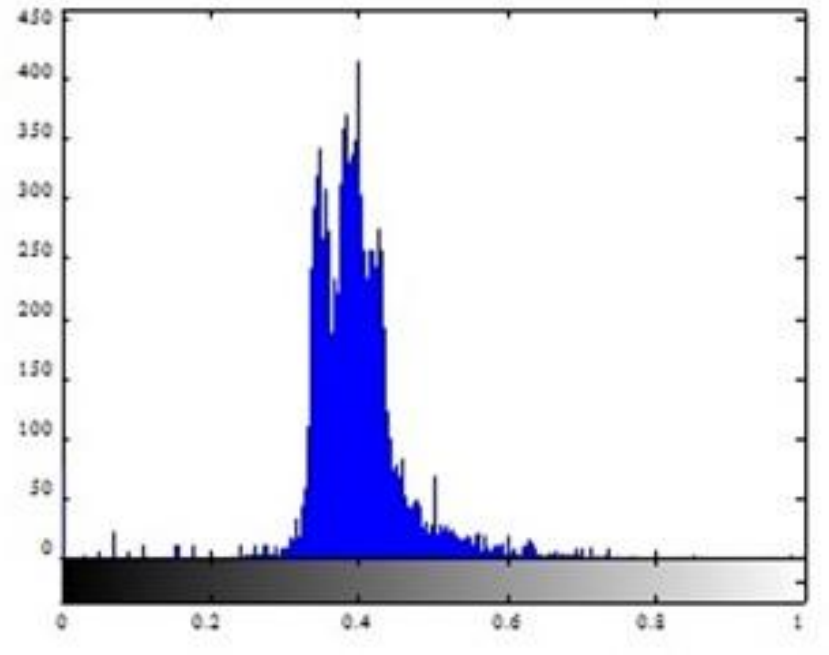


Figure 2.4: Histogram of Normalized R component for the skin color.

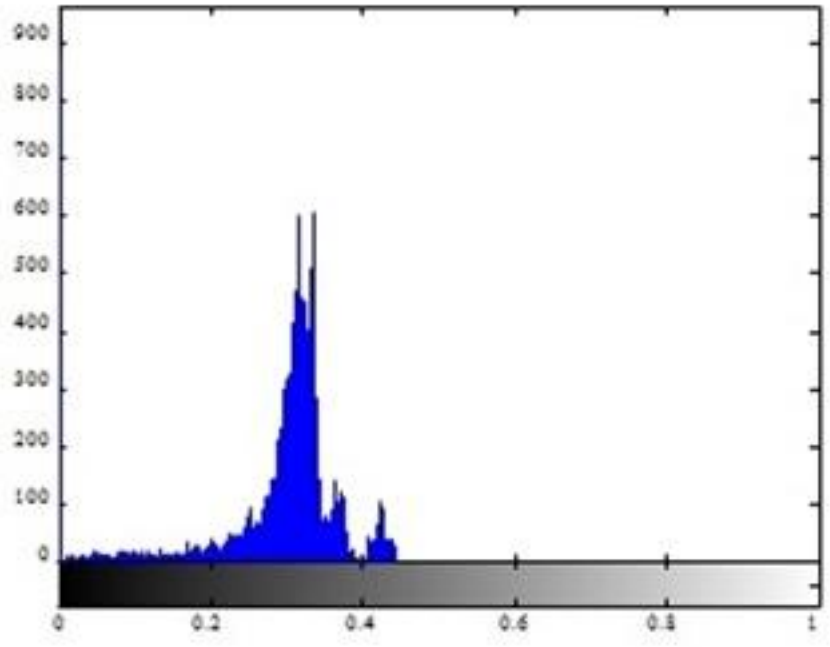


Figure 2.5: Histogram of Normalized G component for the skin color.

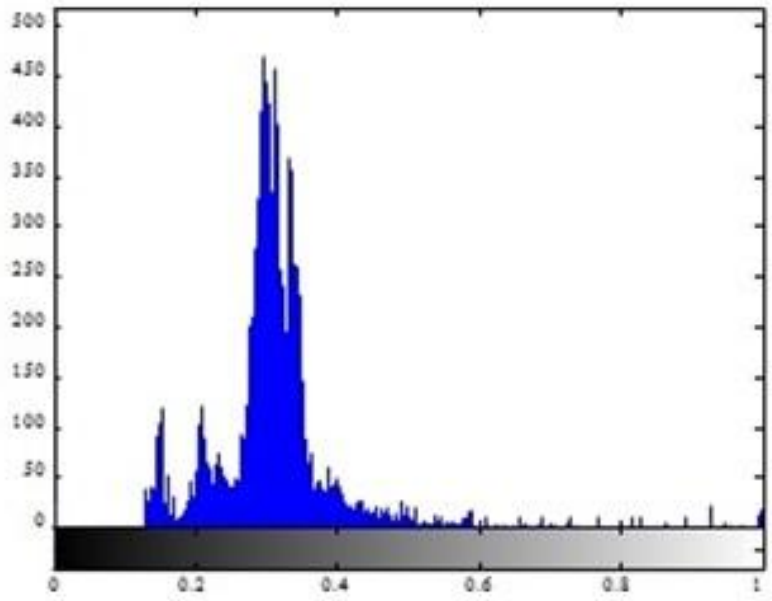


Figure 2.6: Histogram of Normalized B component for the skin color.

2.3.3 YCbCr COLOR SPACE:

The YCbCr is another color space unlike the RGB color space, here the luminance or brightness or intensity is separated from the chrominance or pure color value. The value of Y represents the luminance value and Cb and Cr represents the color or chrominance value, these are also known as color difference of the image. The advantage of this color space is it reduces the computational complexity in image processing. So from the above discussion, we conclude that in YCbCr color space the brightness component is decoupled from the color component of the image which is not possible in the RGB color space and Normalized RGB color space.

$$Y = C1 * R + C2 * G + C3 * B \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.13)$$

$$Cb = (B - Y)/(2 - 2 * C3) \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.14)$$

$$Cr = (R - Y)/(2 - 2 * C1) \quad \ldots\ldots\ldots\ldots\ldots\ldots (2.15)$$

Where $C1$,$C2$ and $C3$ are the standard values for an image and they varies with the quality of the image.

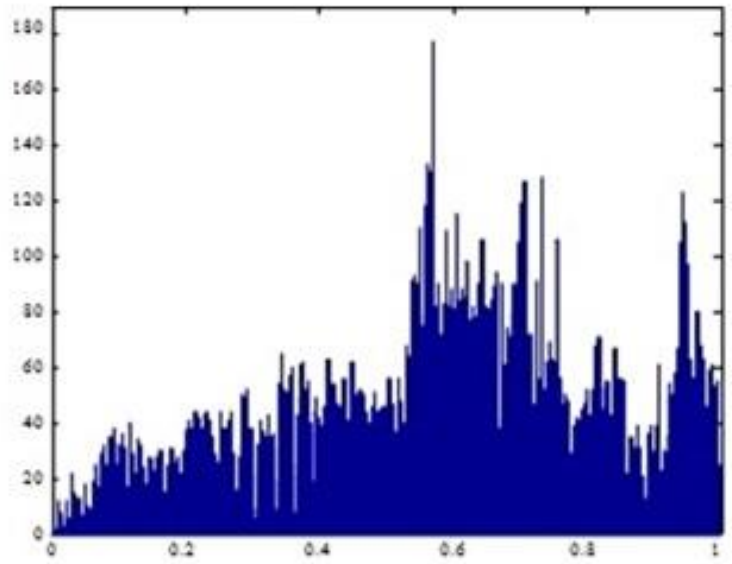


Figure 2.7: Histogram of Y component for the skin color.

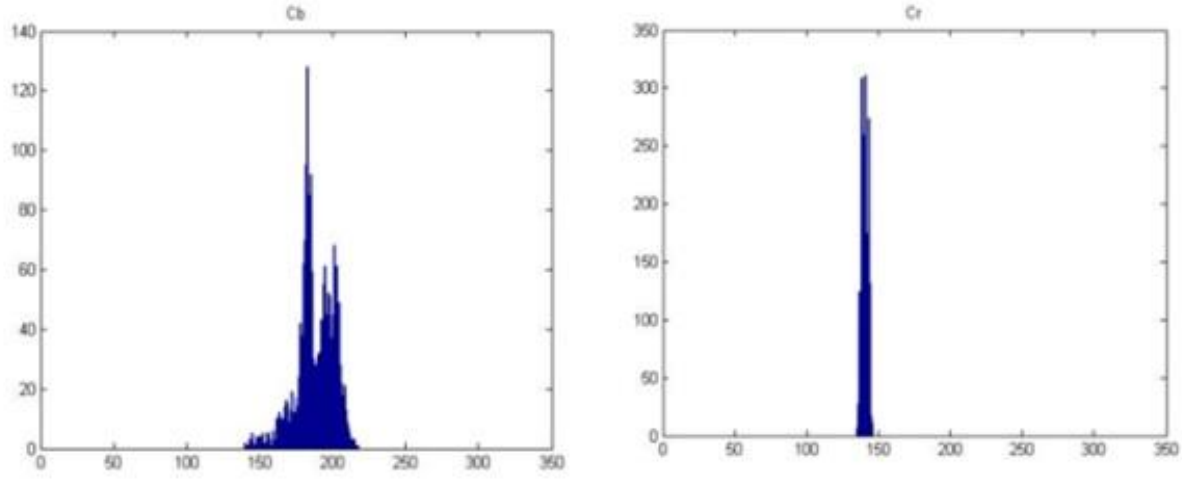


Fig 2.8: Histogram of Cb and Cr of skin color.

2.3.4 HSI COLOR SPACE:

The HSI (hue, saturation and intensity) is similar to HSV model. The main difference between these two models is the computing of the brightness component (I and V), which determines the distribution and dynamic range of both the brightness and saturation. The HSI method is best color space for the traditional image processing function like Convolution, Equalization, and Histogram.

The HSV (hue, saturation, value) color space is developed to be more intuitive in manipulating color and designed to approximate the way humans perceive and interpret color. The HSV color space is preferred for manipulation of hue and saturation i.e. to shift color or adjust the amount of color since it yields a greater dynamic range of saturation.

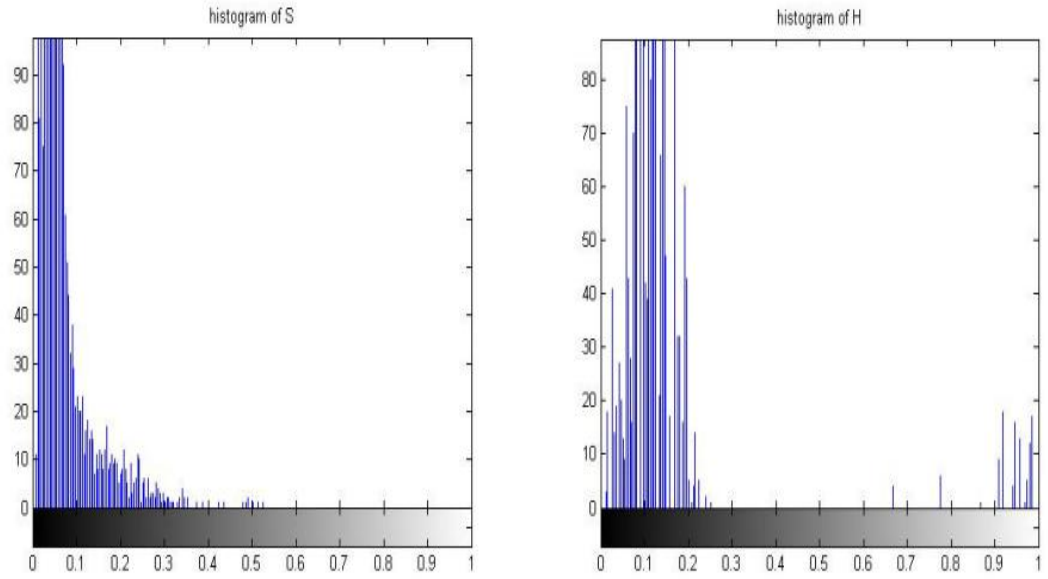


Fig 2.9: Histogram of H and S of skin color.

The range of H and S for segmentation is 0.01 to 0.43 and 0.001 to 0.278. The output pixel at point (x, y) is classified as skin color and set to one if H, S, and value at that point falls inside their respective ranges R_H, R_S values . Other-wise, the pixel is classified as non-skin color and set to zero.

CHAPTER- 3:

HAND GESTURE SEGMENTATION FOR IMAGES

3.1 OVERVIEW OF SEGMENTATION METHOD:

The aim is to automatically segment the hand gesture from a given image under different luminance conditions and complex backgrounds. The luminance value affects the color component of an image which leads to increase the noise level in the segmented image. This work proposes a combined model of two color spaces i.e., HSI, YCbCr and morphological operations with labeling to improve the segmentation performance of color hand gesture from complex backgrounds in terms of completeness and correctness. The proposed color model separates the chrominance and luminance components of the image. The performance of the proposed method is demonstrated through simulation and the experimental results reveal that proposed method provides better performance accuracy compared to the HSI and YCbCr methods individually in terms of correctness and completeness.

3.2 INTRODUCTION:

Hand gesture recognition is an essential task of today's research due to the increasing demands for human–computer interactions (HCIs) in recent years. Image segmentation is the process of partitioning an image into its constituent components i.e. homogeneous and meaningful regions acceding to their identical set of properties or attributes. Segmentation algorithms are based on different parameters of an image like gray-level, color, texture, depth or motion. The image segmentation process can be considered as one of the basic and important steps in digital image processing and computer vision applications such as tracking, pattern recognition and object identification. It is easy to distinguish the objects from the simple back ground but extraction of objects from the complex background of a digital image has been a challenging task in the field of digital image processing. With the increasing demand for complex image analysis and interpretation, the demand for accurate segmentation of images has also grown stronger and as a result many image segmentation methods and algorithms have been developed over the past few decades. The most popular method to perform image segmentation is gray level segmentation method which is based on thresholding because it is simple and having a high speed of operation and ease of implementation.

However the disadvantage of thresholding method is performance limited and suitable for only simple background images. All the color spaces are mathematical representation of a set of colors. All the color spaces are derived from the RGB information supplied by devices such as cameras and scanners. The most common are YCbCr, HSV, HSI, color spaces. The HSV (hue, saturation, value) color space is developed to be more intuitive in manipulating color and designed to approximate the way humans perceive and interpret color. The HSV color space is preferred for manipulation of hue and saturation i.e. to shift color or adjust the amount of color since it yields a greater dynamic range of saturation.

The HSI (hue, saturation and intensity) is similar to HSV model. The main difference between these two models is the computing of the brightness component (I and V), which determines the distribution and dynamic range of both the brightness and saturation. The HSI method is best color space for the traditional image processing function like Convolution, Equalization, and Histogram. The YCbCr is another color space unlike the RGB color space, here the luminance or brightness or intensity is separated from the chrominance or pure color value. The value of Y represents the luminance value and Cb and Cr represents the color or chrominance value, these are also known as color difference of the image.

In this work, a combinational method of HSI and YCbCr is proposed to improve the segmentation performance. In the proposed model, brightness and luminance components are decoupled from the color information of the image which is not possible in the RGB color space. In this proposed method, the segmentation process is carried out by taking the Cb, Cr and H, S values into consideration and morphological operations with labeling is done to improve the performance of segmentation [11-16]. This method uses unsupervised segmentation algorithm, hence no manual adjustment of any design parameter is needed in order to suit any particular input image. Moreover, the algorithm can be implemented in real time, and its underlying assumptions are minimal.

3.3 DATABASE:

In this work, the proposed segmentation model is tested on subset of complex background color hand gesture images of standard database. This data base consists of 60 complex background images of 15 hand postures performed by 4 persons under different luminance conditions [4].

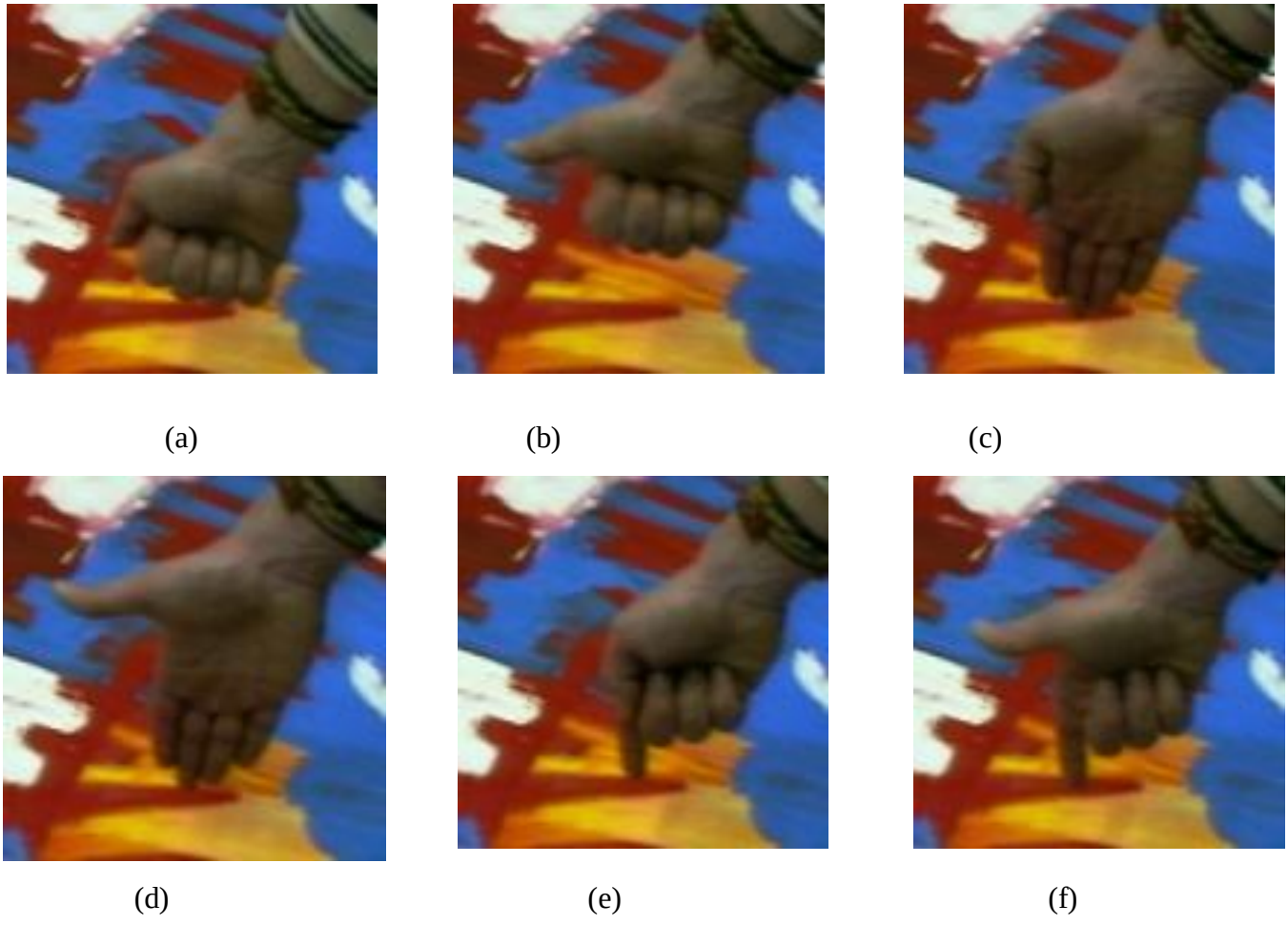


Fig 3.1: Different Hand gesture images for complex background.

3.4 PROPOSED METHODOLOGY:

In this work, a hybrid method is proposed by combining HSI and YCbCr color models to extract the hand gesture from the images with complex background. The outline of the proposed method is as follows.

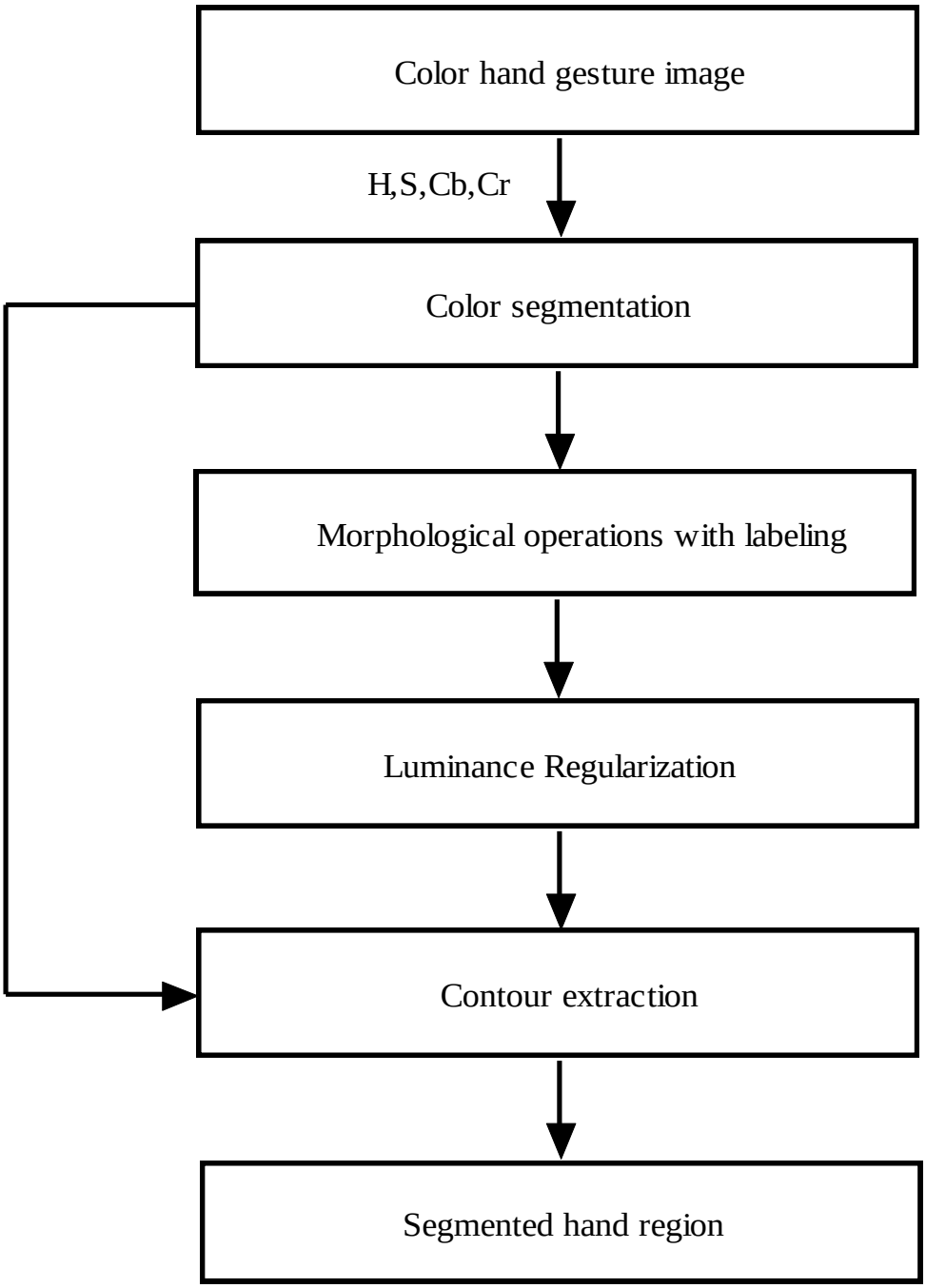


Fig 3.2: Proposed hybrid model.

3.4.1 COLOR SEGMENTATION:

The aim of this step is to divide the image into two classes one is skin color and another is non skin color. In this step, the color segmentation process is carried out by taking Cb, Cr and H, S values in to consideration. For this purpose, the skin color distribution in the YCbCr and HSI

color space is derived [12]. By using the histogram method we have found that a skin-color region can be identified by the presence of a certain set of chrominance (i.e, Cr and Cb) values narrowly and consistently distributed in the YCbCr color space. The location of these chrominance values have been found and can be illustrated using the histogram of the skin color diagram which is shown in below Fig 3.3

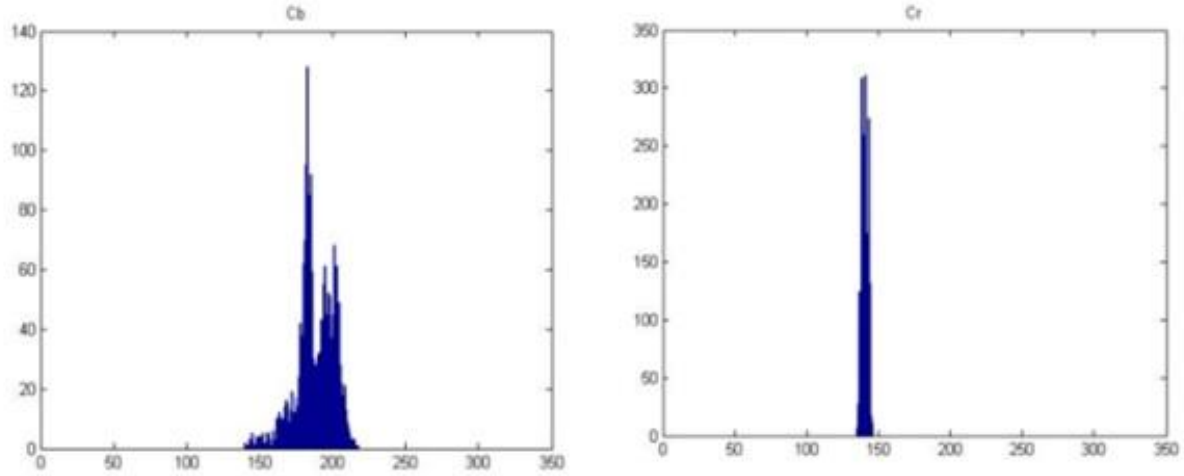
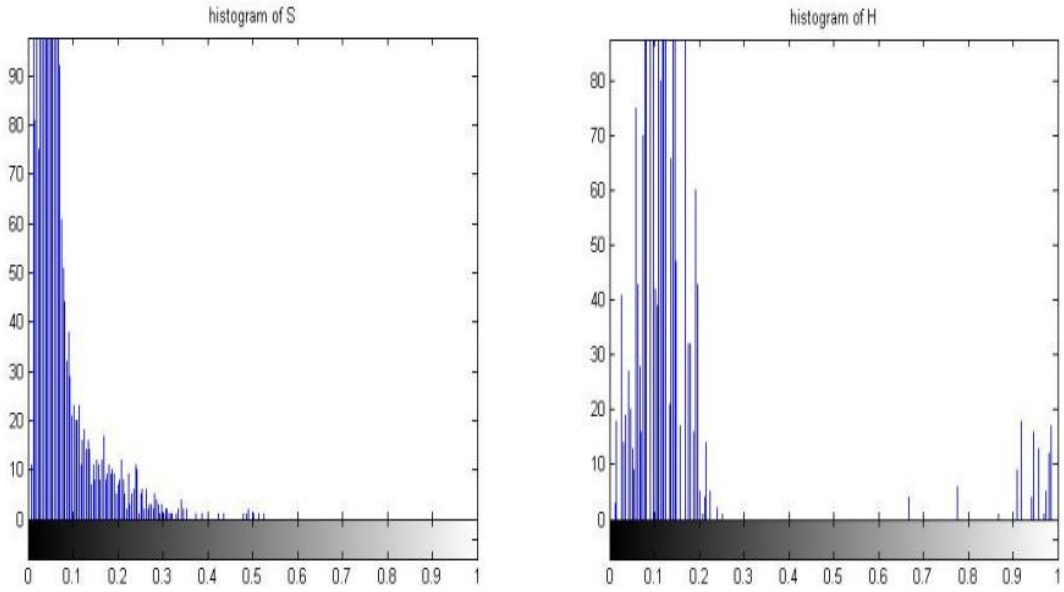


Fig 3.3: Histogram of Cb and Cr of skin color.

The threshold value for the HSI method is calculated using the histogram method. By using histogram method we have been found that range of H and S for skin color is 0.01 to 0.43 and 0.001 to 0.278. The histogram of H, S for different skin colors is shown in below Fig 3.4.



(a) (b)

Fig 3.4: Histogram of H and S of skin color.

The range of H and S for segmentation is 0.01 to 0.43 and 0.001 to 0.278 and for Cb and Cr is 118 to 255 and 128 to 165 which are constant for database system. The output pixel at point (x, y) is classified as skin color and set to one if H, S, Cr and Cb value at that point falls inside their respective ranges $R_H, R_S, R_{cb},$ and R_{cr} . Other-wise, the pixel is classified as non-skin color and set to zero. The segmentation equation can be written as

$$o_1(x,y) = 1, if\,[H(x,y) \in R_H] \cap [S(x,y) \in R_S] \cap [C_r(x,y) \in R_{cr}] \cap [C_b(x,y) \in R_{cb}]$$

0, otherwise. …….. (3.1).

where $x = 0, 1, 2 \ldots, \left(\frac{M}{4} - 1\right) and\ y = 0,\ 1, 2 \ldots, (\frac{N}{4} - 1)$. $R_H, R_S, R_{cb},$ and R_{cr} represents the respective histogram skin color ranges of H, S, Cb, and Cr respectively. The output resolution of the segmented bitmap image in this stage is same as input image i.e. $M \times N$.

3.4.2 MORPHOLOGICAL OPERATIONS WITH LABELING:

The main purpose of this step is to reduce the noise in the image using erosion and dilation process. The density map for the proposed model can be written as

$$D(x,y) = \sum_{i=0}^{3} \sum_{j=0}^{3} o_1(4x+i, 4y+j) \ldots\ldots\ldots \text{(3.2).}$$

where $x = 0, 1, 2 \ldots, \left(\frac{M}{4} - 1\right) and\ y = 0, 1, 2 \ldots, (\frac{N}{4} - 1)$. The equation (4.2) gives the density value of a 4×4 window of the image and its value lies within the range of 0 to 16. This density map is divided in to 3 parts, full density, intermediate density and zero density points. A group of points with zero density value will represent a non-skin region, while a group of full density points will signify a cluster of skin-color pixels and a high probability of belonging to a skin region. Any point of intermediate density value will indicate the presence of noise. Once the density map is derived, the process termed as density regularization starts.

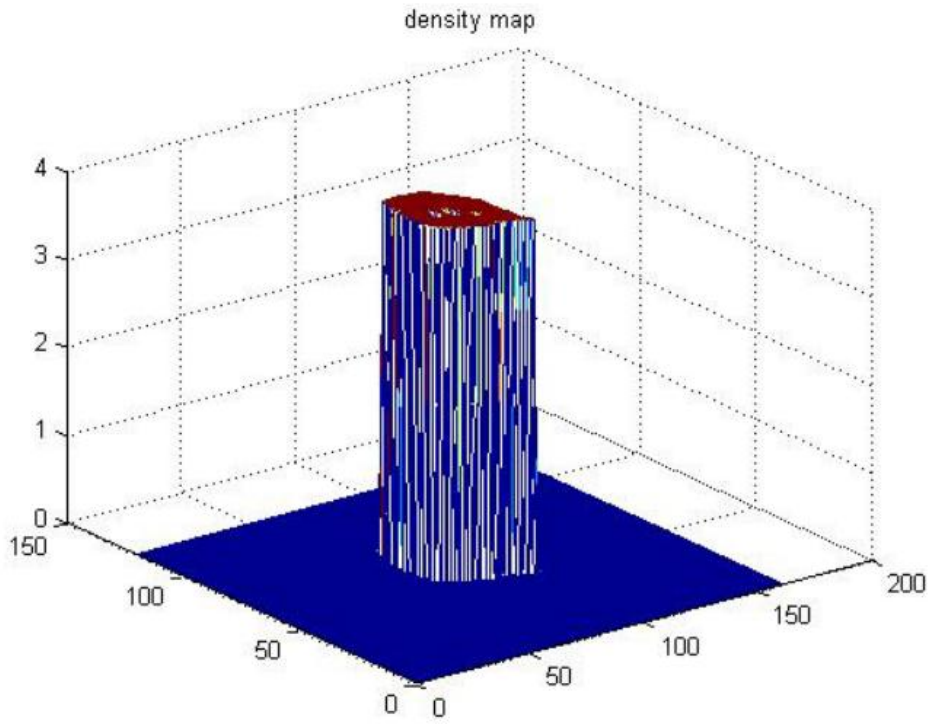


Fig 3.5: Density map of a hand image.

After finding the density map of the image the four steps of operation for density regularization is performed. Density regularization involves the following four steps:

• Discard all points at the boundary or edge of the density map, i.e., set D(0, y)=D($\frac{M}{4}-1$, y) =D(x,0)= D(x,$\frac{N}{4}-1$)=0. for all $x = 0, 1, 2 \dots, \left(\frac{M}{4}-1\right) and\ y = 0,1, 2 \dots, (\frac{N}{4}-1)$.

• Erode any full-density point (i.e., the points having value 16 set to zero) if it is surrounded by less than five other full-density points in its local 3 × 3 neighborhood.

• Labeling works by scanning an image pixel by pixel (from top to bottom and left to right) in order to identify connected pixels. It removes all the minimum connected parts in image (i.e., noise) which is unwanted and maximum connected part will present in the image.

• Dilate any point of either zero or intermediate density (i.e., set to 16) if there are more than two full-density points in its local 3 × 3 neighborhood.

Then the density map is converted to bitmap using the below equation 3.3

$$o_2(x,y) = \begin{cases} 1, if D(x,y) = 16 \\ 0, \text{otherwise} \end{cases} \quad \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \quad (3.3)$$

for all $x = 0, 1, 2 \dots, \left(\frac{M}{4}-1\right) and\ y = 0, 1, 2 \dots, (\frac{N}{4}-1)$.

3.4.3 LUMINANCE REGULARIZATION:

In complex background images, the brightness is non uniform throughout the skin region, while the background region tends to have a more even distribution of brightness [14]. Hence, in this stage based on these characteristics, background region that was previously detected due to its skin-color appearance can be further eliminated. The analysis employed in this stage involves the spatial distribution characteristic of the luminance values since they define the brightness of the image. In this proposed method, it uses standard deviation as the statistical measure of the distribution. The size of the previously obtained bitmap of $o_2(x,y)$ is (M/4 × N/4). Hence the

each pixel of the image will have information of 4 ×4 window. 'W' is the intensity value represented by the each pixel of the image and its standard deviation of its corresponding group of luminance values can be written as

$$\sigma(x,y) = \sqrt{E[W^2] - (E[W])^2)} \ldots\ldots\ldots \text{(3.4).}$$

The bitmap equation for stage 3 can be written as

$$o_3(x,y) = \begin{cases} 1, if\, o_2(x,y) = 1\ and\ \sigma(x,y) \geq 2 \\ 0, \text{otherwise} \end{cases} \ldots. \text{(3.5).}$$

for all $x = 0, 1, 2 \ldots, \left(\frac{M}{4} - 1\right) and\ y = 0, 1, 2 \ldots, (\frac{N}{4} - 1)$.

3.4.4 CONTOUR EXTRACTION:

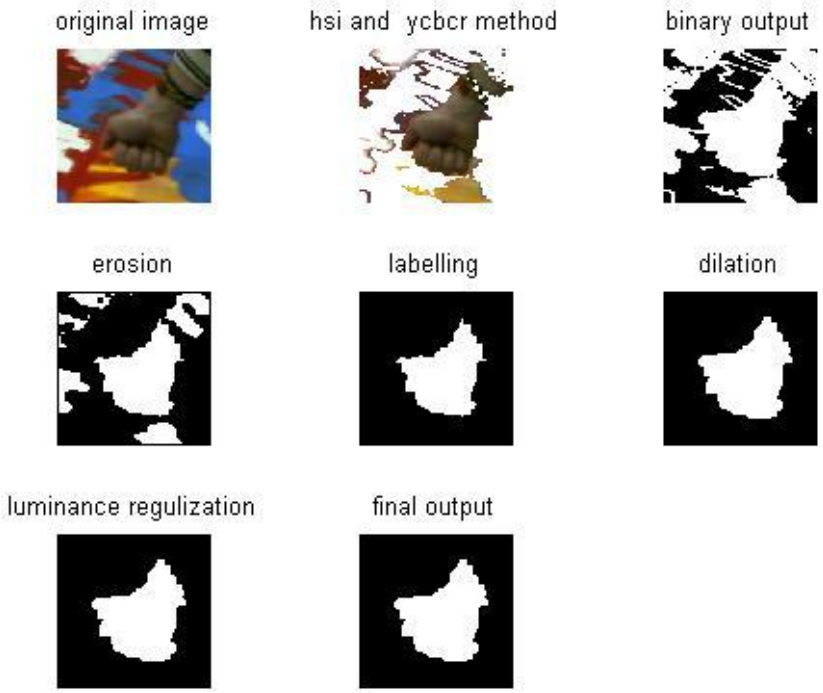
In this final step, the (M/4 × N/4) output bitmap of step four is converted back to the dimension of (M × N). To achieve the increase in spatial resolution, this work utilizes the edge information that is already made available by the color segmentation in step one. Therefore, all the boundary points in the previous bitmap is being mapped into the corresponding group of 4 × 4 pixels with the value of each pixel as defined in the output bitmap of step one [17].

3.5 SIMULATION RESULTS AND DISCUSSION:

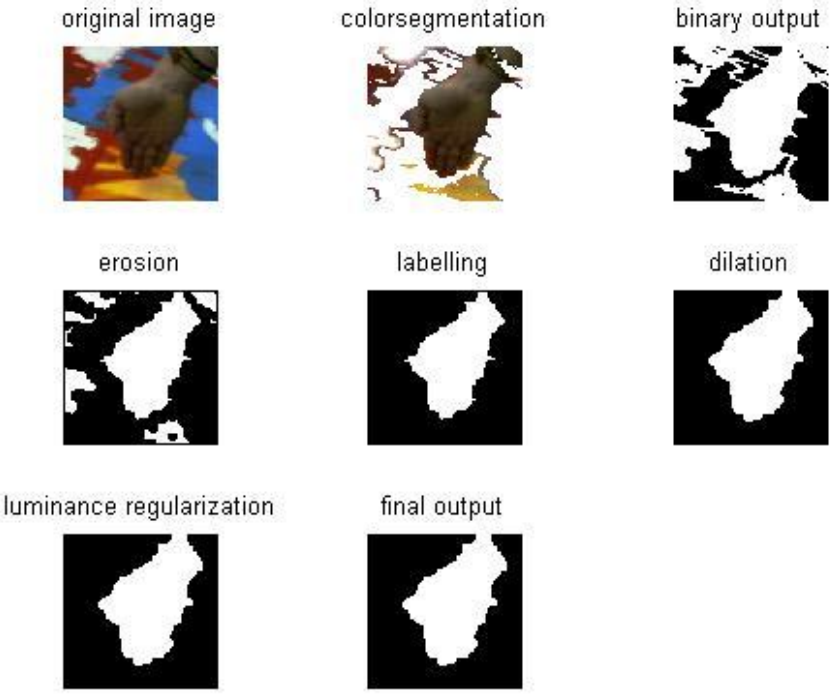
The step wise output of the proposed method is shown in below Fig 3.6.The performance of the proposed method is compared with HSI and YCbCr methods in terms of completeness and correctness [18-20]. The definition of completeness and correctness are described below.

• Completeness is the percentage of the ground truth region extracted by the segmentation algorithm.

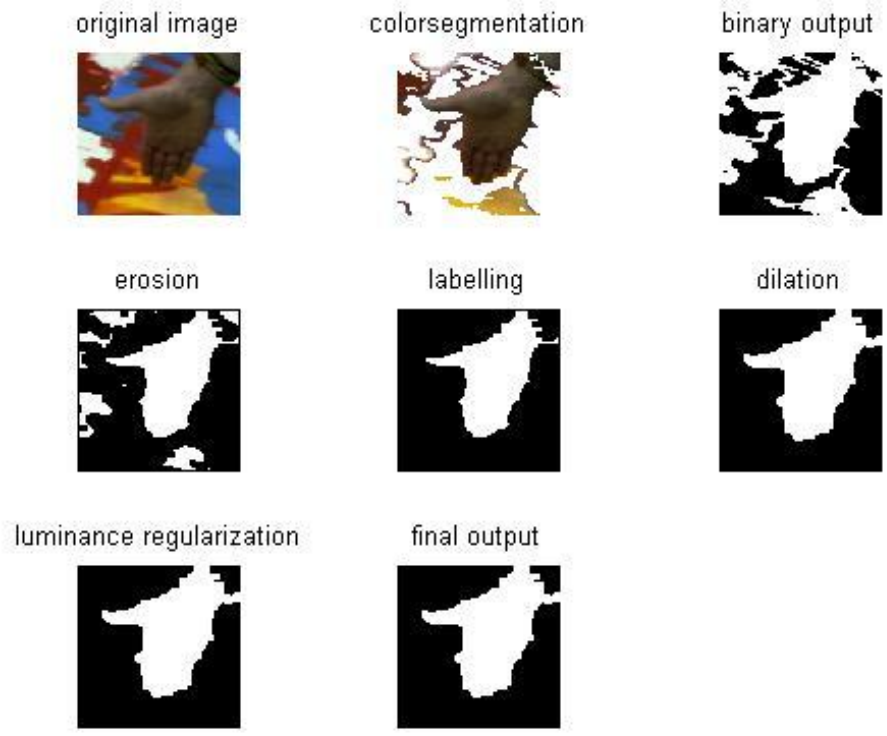
• Correctness can be defined as the percentage of correctly extracted region (ground truth) by the segmentation algorithm.


original image
hsi and ycbcr method
binary output
erosion
labelling
dilation
luminance regulization
final output

(a)


original image
colorsegmentation
binary output
erosion
labelling
dilation
luminance regularization
final output

(b)



(c)

Fig 3.6: Step-wise output of Proposed method for three hand gesture images.

The correctness and the completeness can be calculated using the following formulas.

$$Correctness = \frac{TP}{TP+FP} \times 100\% \quad \text{..... (3.6)}$$

$$Completeness = \frac{TP}{TP+FN} \times 100\% \quad \text{..... (3.7)}$$

- TP (True Positive image) is the common part of Ground truth image and segmented image.
- FN (False Negative image) difference between ground truth image and true positive image.
- FP (False Positive image) difference between segmented image and true positive image.

Here the ground truth image represents the actual output. In this work, the ground truth image is manually created by using *Gimp software*.

Fig 3.7: Segmented images and the corresponding Ground truth images of proposed model.

The ground truth images and corresponding output of the proposed hybrid model is shown in Fig 3.7. Table 2 shows the comparative performance of the proposed segmentation algorithm in terms of correctness and completeness withrespect to HSI and YCbCr methods. The proposed method yields 90.47% of correctness and 98.77% of completeness whereas HSI method provides 84.54% and 93.41% and YCbCr method provides 86.30% and 96.55% respectively. Therefore, the proposed method provides better performance for the color segmentation as compared to the earlier reported HSI and YCbCr methods.

TABLE II: COMPARISON RESULT FOR 60 IMAGES.

Method	Correctness in %	Completeness in%
HSI	84.54	93.41
YCbCr	86.30	96.55
Proposed Model	90.47	98.77

.

3.6 SUMMARY:

In this work, a combinational method of HSI and YCbCr along with the morphological operations with labeling is proposed for background noise reduction and better performance of segmentation. In this work the hand gesture image is segmented by performing region segmentation using skin color map. This is feasible because the skin color distribution is different from background color distribution. By using skin color map, this work classifies pixel of the input image into skin color and non-skin color. Experimental results reveal that the proposed method provides better performance accuracy compared to HSI and YCbCr methods in terms of correctness and completeness. This technique can be used in the field like bio-medical imaging, satellite imaging and robotics, face detection, gesture recognition where accuracy has the higher priority over complexity.

CHAPTER- 4:

DETECTION AND TRACKING OF DYNAMIC HAND GESTURE

4.1 INTRODUCTION:

Tracking and detection of an object is essential task for human computer interface used in many applications. Different imaging techniques for tracking and identification have been proposed by many researchers [21]. Many researchers have their own methods to solve the problem of object detection, object tracking and object identification. In object detection methodology, many researchers have developed their methods. (Liu et.al, 2001) [22] Proposed background subtraction to detect moving regions in an image by taking the difference between current and reference background image in a pixel-by-pixel. It is extremely sensitive to change in dynamic scenes derived from lighting and extraneous events etc. In another work, (Stauffer & Grimson, 1997) [23] proposed a Gaussian mixture model based on background model to detect the object. (Lipton et al., 1998) proposed frame difference that use of the pixel-wise differences between two frame images to extract the moving regions. This method is very adaptive to dynamic environments, but generally does a poor job of extracting all the relevant pixels, e.g., there may be holes left inside moving entities. In order to overcome disadvantage of two-frames differencing, in some cases three-frames differencing is used. For instance, (Collins et al., 2000) [32] developed a hybrid method that combines three-frame differencing with an adaptive background subtraction model for their VSAM (Video Surveillance and Monitoring) project. The hybrid algorithm successfully segments moving regions in video without the defects of temporal differencing and background subtraction. (Desa & Salih, 2004) [28] Proposed a combination of background subtraction and frame difference that improved the previous results of background subtraction and frame difference.

In object tracking methodology, regarding to our study, this article will describe more about the region based tracking. Region-based tracking algorithms track objects according to variations of the image regions corresponding to the moving objects [24-26]. For these algorithms, the background image is maintained dynamically and motion regions are usually detected by subtracting the background from the current image. (Wren et al., 1997) [27] Explored the use of small blob features to track a single human in an indoor environment. In their work, a human body is considered as a combination of some blobs respectively representing various body parts such as head, torso and the four limbs. The pixels belonging to the human body are assigned to

the different body part's blobs. By tracking each small blob, the moving human is successfully tracked. (McKenna et al., 2000) [29]Proposed an adaptive background subtraction method in which color and gradient information are combined to cope with shadows and unreliable color cues in motion segmentation. Tracking is then performed at three levels of abstraction: regions, people, and groups. Each region has a bounding box and regions can merge and split. A human is composed of one or more regions grouped together under the condition of geometric structure constraints on the human body, and a human group consists of one or more people grouped together. Moreover, for object identification, (Cheng & Chen, 2006) [30]proposed a color and a spatial feature of the object to identify the track object. The spatial feature is extracted from the bounding box of the object. Meanwhile, the color features extracted is mean and standard value of each object. (Czyz et al., 2007) proposed the color distribution of the object as observation model. The similarity of the object is measure using Bhattacharya distance. The low Bhattacharya distance corresponds to the high similarity. To overcome the related problem described above, this article proposed a new technique for object detection employing frame difference on low resolution image (Sugandi et al., 2007), object tracking employing block matching algorithm based on PISC image (Satoh et al., 2001) and object identification employing color and spatial information of the tracked object (Cheng & Chen, 2006).

Generally all tracking systems begin with detection and ends with identification. In the first stage of detection it separates the wanted region from the background of the video sequence. It uses some pre-processing steps like gray scaling, smoothing, morphological operations like erosion and dilation with labeling [31]. These pre-processing steps are applied to reduce the noise, in such a way to increase the accuracy of the tracking.

In the second stage, it involves the tracking of the desired object. In this stage it performs a block matching technique only to track the desired object from the moving objects which are presented in the background. In the third stage, it involves object identification. In this final stage it uses the features which are extracted from the image for accurate identification like color and spatial features. If any new objects appears on the screen it will check for the closeness of the features which are already made available. If the features are matching then it is a desired object, otherwise it is an unwanted object.

4.2 BASIC METHODOLOGY:

In this algorithm it involves three basic steps:

1.) Hand detection. 2.) Hand tracking. 3.) Hand identification.

Basic flow chart of the video tracking is as shown in Fig 4.1.

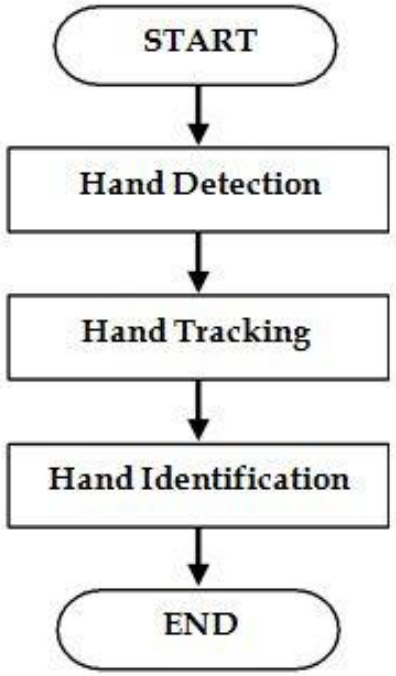


Fig 4.1: Flowchart of tracking a video sequence.

4.3 HAND DETECTION:

In the first step of hand detection, it is mainly divided into two basic steps:

1.) Pre-processing. 2.) Filtering.

4.3.1 PRE-PROCESSING:

The first step on the moving object detection process is capturing the image information using a video camera. Image is capture by a video camera as 24 bit RGB (red, green, blue) image which

each color is specified using 8-bit unsigned integers (0 through 255) that representing the intensities of each color [32]. The size of the captured image is 320x240 pixels. This RGB image is used as input image for the next stage. In order to reduce the processing time, gray-scale image is used on entire process instead of color image. The gray-scale image only has one color channel that consists of 8 bit while RGB image has three color channels.

Image smoothing is performed to reduce image noise from input image in order to achieve high accuracy for detecting the moving objects. The smoothing process is performed by using a median filter with m×m pixels. The un-stationary background often considers as a fake motion other than the motion of the object interest and can cause the failure of detection of the object. To handle this problem, we reduce the resolution of the image to be a low resolution image. A low resolution image is done by reducing spatial resolution of the image with keeping the image size (Gonzales & Woods, 2001) and (Sugandi et al., 2007).

In this work, the low resolution image is done by averaging pixels value of its neighbors, including itself. We use a video image with resolution 320x240 pixels. The original image size is 320x240 pixels. After applying the low resolution image, the numbers of pixels will be 160x120, 80x60, or 40x30 pixels, respectively, while the image size is still 320x240 pixels. The low resolution image can be used for reducing the scattering noise and the small fake motion in the background because of un-stationary background. These noises that have small motion region will be disappeared in low resolution image.

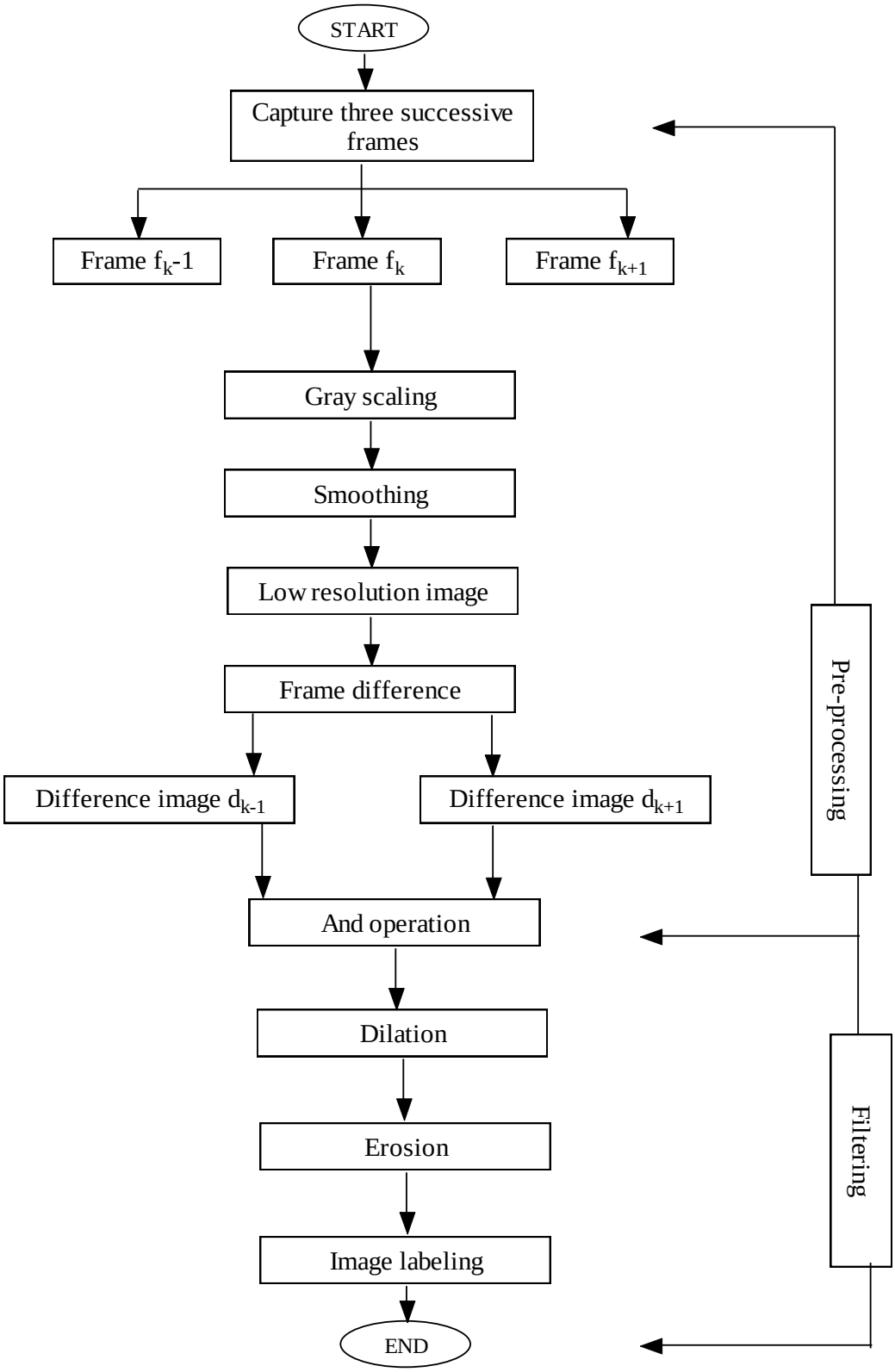


Fig 4.2: Flow chart of hand detection [21].

4.3.2 FILTERING:

In the second stage filtering is done by applying morphological operations such as erosion and dilation along with the labeling. According to the density value, we classify each point into three types, namely, zero (D=0), intermediate (0< D < 16), and full (D=16). A group of points with zero density value will represent a non-skin region, while a group of full density points will signify a cluster of skin-color pixels and a high probability of belonging to a skin region. Any point of intermediate density value will indicate the presence of noise.

• Discard all points at the boundary or edge of the density map, i.e., set D(0,y)=D(M/8-1, y) =D(x,0)= D(x, N/8-1)=0 for all x=0....M/8-1 and y=0.....,N/8-1.

• Erode any full-density point (i.e., the points having value 16 set to zero) if it is surrounded by less than five other full-density points in its local 3 x 3 neighborhood.

• Dilate any point of either zero or intermediate density (i.e., set to 16) if there are more than two full-density points in its local 3 x 3 neighborhood.

Labeling works by scanning an image pixel by pixel (from top to bottom and left to right) in order to identify connected pixels. It removes all the minimum connected parts in image (i.e., noise) which is unwanted and maximum connected part will present in the image.

In this step, frame difference is taken between three successive frames i.e. between f_k and f_{k-1} and between frame f_k and f_{k+1}. The output image as frame difference image is two difference images d_{k-1} and d_{k+1}. Threshold is performed by threshold value T on the difference image d_{k-1} and$d_{k+1}$to distinguish between background and desired object of the image. The process is followed by applying AND operation to d_{k-1} and d_{k+1}. The output image of this operation is named as motion masking to reduce the noise as much as possible.

4.4 HAND TRACKING:

4.4.1 BLOCK MATCHING TECHNIQUE:

After the hand detection, we have to go to next step is hand tracking [33-36]. Tracking process can be considered as a region mask association between temporally consecutive frames and

estimating the trajectory of an object in the image plane as it moves around a scene. For this purpose we use Block matching technique. Block matching is a technique for tracking the interest moving object among the moving objects emerging in the scene. In this step, the blocks are defined by dividing the image frame into non-overlapping square parts. The blocks are made based on peripheral increment sign correlation (PISC) image (Satoh et al., 2001; Sugandi et al., 2007) that considers the brightness change in all the pixels of the blocks relative to the considered pixel.

To determine the matching criteria of the blocks in two successive frames, we evaluate using correlation value. To determine the closeness i.e. matching criteria between the two successive frames we use correlation between two frames which gives us clear idea of how far they are matching. The high correlation value shows that the blocks are matched each other. The interested moving object is determined when the number of matching blocks in the previous and current frame is higher than the certain threshold value.

4.4.2 TRACKING METHOD:

Tracking has been approached in two basic image processing frameworks. One is to track image features, that is, image regions with special properties making them clearly identifiable and efficiently detectable. Classic examples of features in computer vision are corners lines and deformable contours. Such features are detected automatically in each frame and tracked through a sequence for as long as possible. The resulting motion field is sparse: all motion information computed refers only to the feature regions. Finding the same feature in subsequent frames is a problem similar to stereo matching [37].The alternative framework is optical flow methods, a class of differential techniques estimating image motion at each pixel. The advantage is that a dense motion field is produced and, in principle, can be used for segmentation, tracking, and 3-D motion analysis. In practice, the main disadvantage with underwater sequences seems to be the differential nature of the methods, which requires a high frame rate and negligible changes between consecutive frames. We make the block size (block *A*) with 9x9 pixels in the previous frame. We assume that the object coming firstly will be tracked as the desired moving object. The block *A* will search the matching block in each block of the current frame by using correlation value. In the current frame, the interest moving object is tracked when the object has maximum number of matching blocks.

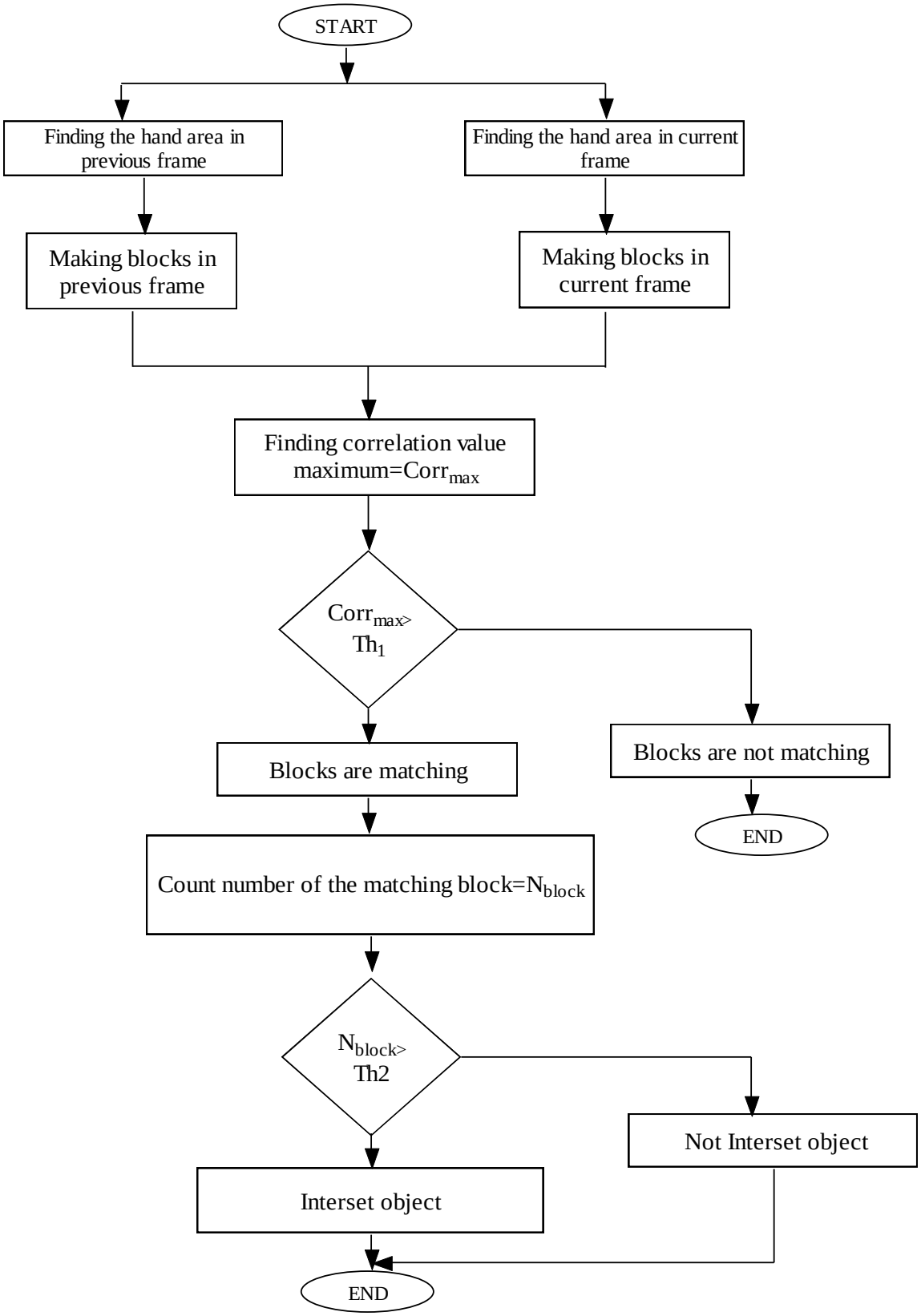


Fig 4.3: Flow chart of hand tracking [21].

4.5 HAND IDENTIFICATION:

The last stage involved in this basic flow chart is hand detection. It can be identified by a set of

Features [38]. Here the feature extraction is done by spatial feature extraction and color feature extraction taking into consideration.

4.5.1 SPATIAL FEATURE EXTRACTION:

The features of the spatial extraction give the position of the tracked object. Then these features are mixed with the time domain features which gives trajectory of the object i.e. start and end positions of the hand so that we can easily estimate the various parameters of the objects like speed etc. which are very important for hand identification. The bounding box is used for the spatial information of the objects. After getting the desired moving object we extract the desired moving object by using bounding box.

4.5.2 COLOR FEATURE EXTRACTION:

The color feature can be extracted from the object which is RGB color space as the RGB color information can be obtained from video capture device directly. The mean value is calculated for each color component of RGB space. Standard deviation is a statistical term that provides a good indication of volatility. It measures how widely the values are dispersed from the average. Dispersion is the difference between the actual value and the average value. The larger the difference between the actual color and the average color is, the higher the standard deviation will be, and the higher the volatility. We can extract more useful color features by computing the dispersed color information.

4.5.3 IDENTIFICATION PROCESS:

After the feature extraction**,** the feature set is saved by as set of vectors. To identify a moving object, a feature queue is created to save the features of the moving objects. When a new object enters the system, it will be tracked and labeled, and the features of the object are extracted and

recorded into the queue. Once a moving object is detected, the system will extract the features of the object and identify it from the identified objects in the queue by computing the similarity.

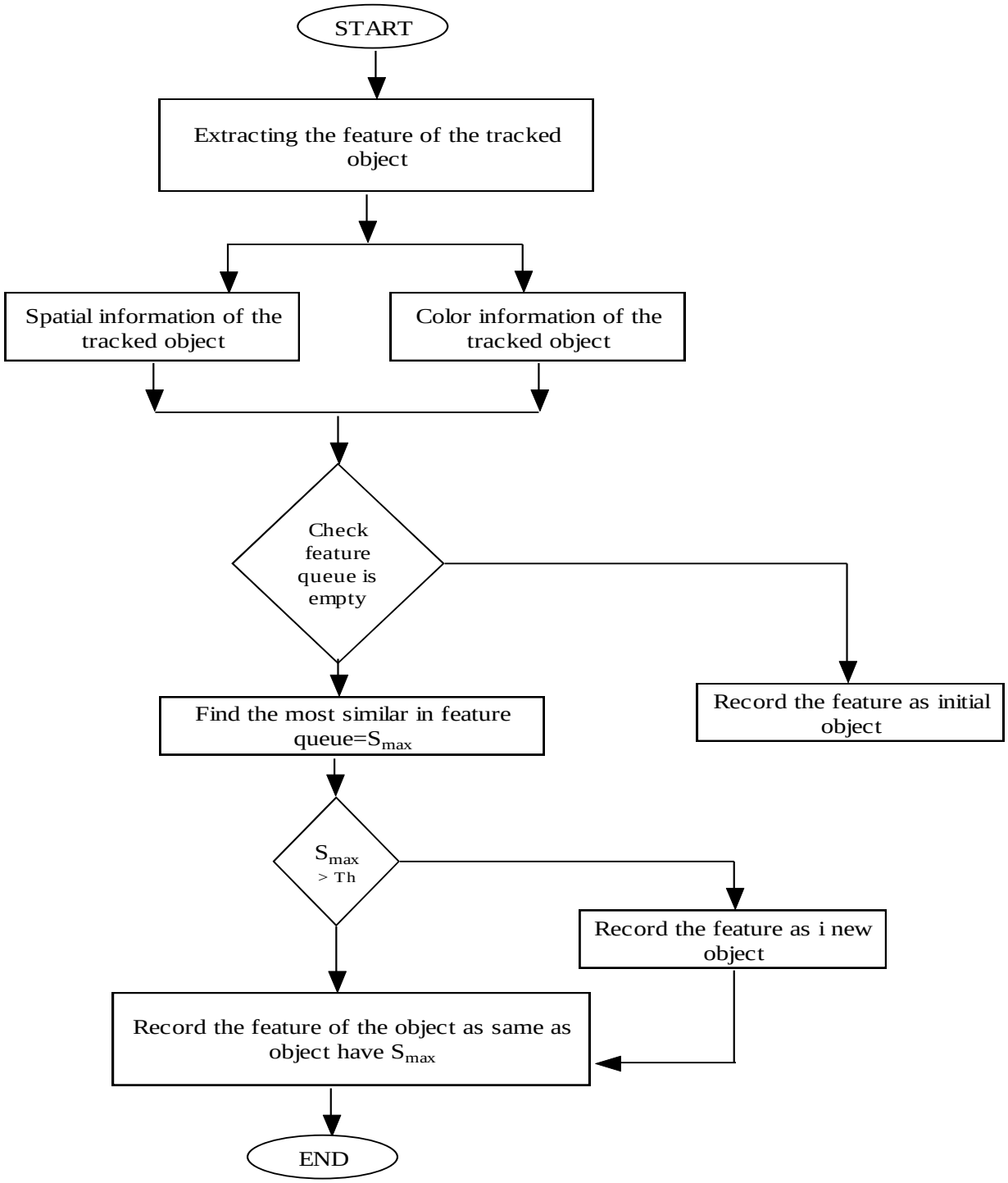
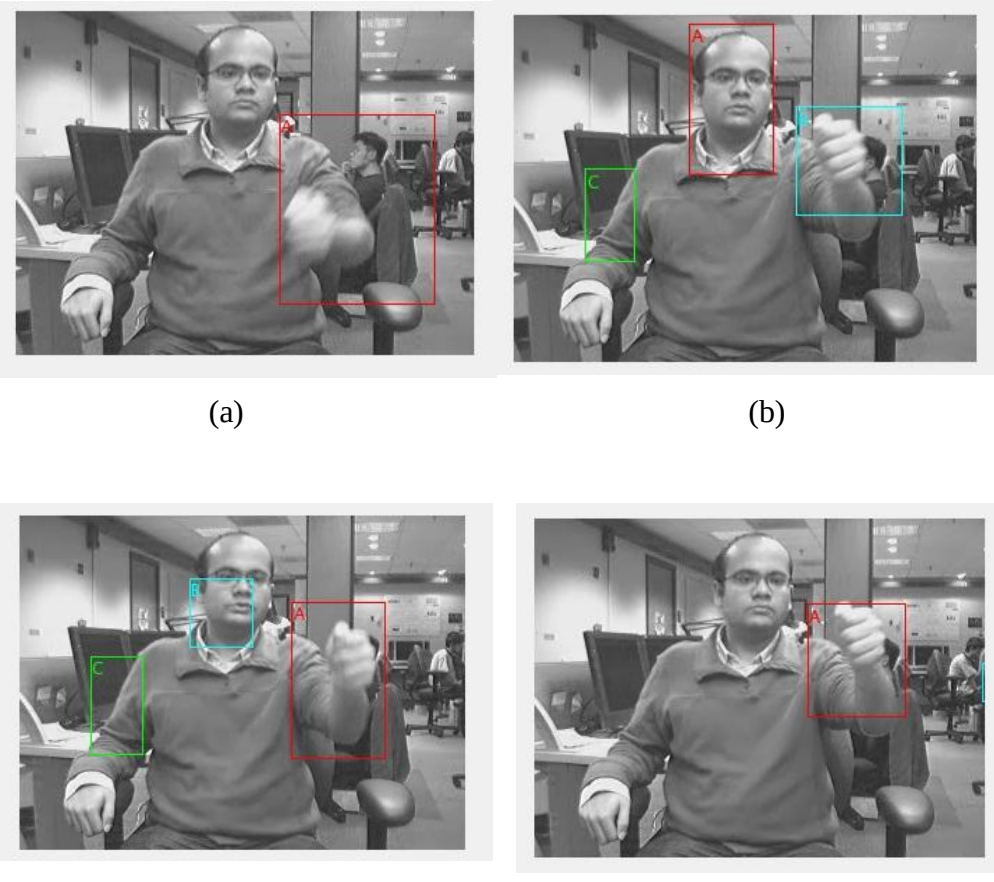


Fig 4.4: Flow chart of hand identification [21].

4.6 SIMULATION RESULTS:

The experiments are conducted in different environment and real time conditions. The experiment is performed in 2.54 [GHz] Pentium 4 PC with 512 MB memory. The image resolution is 420× 240 pixels. The size of each block is 8 × 8 pixels [39-41]. The experimental results are shown in Fig. 4.5. The rectangle area on the object shows the tracked object. The identification result is shown in Table 3.In this work all the operations are performed on video sequence. The standard database which we used consists of 3persons performing in different environments is available in online with different luminance or brightness conditions. In the experimental results, we can extract the moving objects on the successive frame successfully and identification rates of 90.8% were achieved



(a) (b)

(c) (d)

Fig 4.5: Simulation results for Hand tracking.

TABLE III: COMPARISON RESULT FOR HAND DETECTION IN DIFFERENT ENIVRONMENTS.

Experiment	Hand detected	Correct identification	Identification rate in %
Easy	**110**	**99**	**90.8**
Moderate	**117**	**100**	**85.47**
Hard	**130**	**106**	**81.53**

In this work, tracking and detection of hand is done by using block matching algorithm for detecting the moving object employing frame difference on low resolution image, tracking the moving object employing block matching technique based on peripheral increment sign correlation image for tracking the interest moving object among the moving objects emerging in the background and identifying the moving objects employing color and spatial information of the tracked object. The experiment results and data show the effectiveness and the satisfaction of the proposed methods.

CHAPTER 5:

CONCLUSIONS

CONCLUSIONS:

In this work, it is divided into two parts i.e. static and dynamic. In the first part the segmentation of hand gesture region is carried out with proposed methodology by combining two color spaces like HSI and YCbCr color spaces along with morphological operations such as erosion and dilation with labeling. In this work the hand gesture image is segmented by performing region segmentation using skin color map. This is feasible because the skin color distribution is different from background color distribution. By using skin color map, this work classifies pixel of the input image into skin color and non-skin color. Experimental results reveal that the proposed method provides better performance accuracy compared to HSI and YCbCr methods in terms of correctness and completeness. This technique can be used in the field like bio-medical imaging, satellite imaging and robotics, face detection, gesture recognition where accuracy has the higher priority over complexity.

The second part is dynamic, tracking and detection of hand is done by using block matching algorithm for detecting the moving object employing frame difference on low resolution image, tracking the moving object employing block matching technique based on peripheral increment sign correlation image for tracking the interest moving object among the moving objects emerging in the background and identifying the moving objects employing color and spatial information of the tracked object. The experiment results and data show the effectiveness and the satisfaction of the proposed methods. However, the proposed method still has limitations. The limitations can be investigated as followings. Firstly, the detection method based on frame difference on low resolution image has a limitation when the moving object is too small to be detected. It is occurred because the low resolution image removes the small moving objects emerging in the background. To overcome this limitation, we can add another method such as skin color detection. By using this method, even if the moving object is too small, it can still be detected based on the skin color of the object. Secondly, the block matching technique has successfully tracked the interest moving object in the occlude condition. However, when the moving objects appear in the same time, we cannot judge any object to be an interest object. Moreover, when the interest moving object is covered by the occluded object, the image information of the interest moving object cannot be read by the camera. This condition cause the system cannot recognize the interest moving object. Those limitations can be solved by adding

other information to the interest moving object such as flow of moving object based on optical flow, dimension or another feature and also we can add the color information to each object. So whenever the objects appear, they have their own model that different from each other. And we can track them based on the model. Thirdly, color and spatial information method show the high correct identification rate.

FUTUREWORK:

However, the system still cannot identify the objects sometimes when they are just entering or leaving the scene. The extracted features in this case are not enough to be used to identify the moving objects. The system also has limitation when the object is moving slowly. In this condition, the inter-frame difference image of the object will become smaller and we will get smaller bounding box and less moving pixels. Therefore, the extracted features will lose its representative. The correct identification rate highly depends on the correctness of the moving object detection and feature representation. This problem can be improved by a better feature selection method and moving object detection method. By considering those limitations and implement some improvements to our method including speed up the processing time, they could lead to some improvements in the tracking system. These are remaining for future work.

PUBLICATIONS:

1. D. Avinash Babu, D. K. Gosh, and S. Ari, “Color Hand Gesture Segmentation for Images with Complex Background”, in *proceedings of 2013 International Conference on Circuits, Power and Computing Technologies (ICCPCT-2013)*, 2013.

REFERENCES:

[1] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, Pearson Education Third Edition, 2009.

[2] P. L. Correia and F. Pereira, "Objective Evaluation of Video Segmentation Quality," *IEEE transactions On Image Processing*, vol. 12, no. 2, Feb. 2003.

[3] D. Chai and K. M. Nagan , "Face segmentation using skin color mapping video phone applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, June 1999.

[4] Database is available online (http://www.idiap.ch/marcel/Databases/ main.html).

[5] G. M. N. R Gajanayake and R. D. Yapa, "Comparison of Standard Image Segmentation Methods for Segmentation of Brain Tumors from 2D MR Images," *International conference on Industrial and Information Systems,* Dec. 2009.

[6] M. Silveir, "Comparison for Segmentation Method for Melanoma Diagnosis in Dermoscopy Images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 1, Feb. 2009.

[7] Y. Zhu, H. Ren, G. Xu, X. Lin, "Toward real-time human – computer interaction with continuous dynamic hand gestures," in *Proc. Fourth Intl.Conf. Automatic. Face Gesture Recognition, Grenoble, France, IEEE Comput. Soc.*, pp. 544 – 549, 2000.

[8] N. Otsu, "A Threshold Selection Method from Gray-Level Histogram," *IEEE Trans. Systems Man, and Cybernetics*, vol.9, pp.62-66, 1979.

[9] C. K. P. Clarke, Color Encoding and Decoding Techniques for Line-Locked Sampled PAL and NTSC Television Signals, BBC Research Department Report BBC RD 1986/2.

[10] M. T. Chan, "HMM-based audio-visual speech recognition integrating geometric and appearance-based visual features," in *IEEE Fourth Workshop on Multimedia Signal Processing, Cannes, France,* October 2001, pp.9–14.

[11] M. Kolsch and M. Turk, "Robust Hand Detection," in *Proc. IEEE Intl. Conference on Automatic Face and Gesture Recognition,* May 2004.

[12]A Thesis in "Hand Gesture Recognition System" by EMRAH GINGIR, Sep 2010.

[13] Aryuanto Soetedjo, Koichi Yamada, Skin Color Segmentation Using Coarse-to-Fine Region on Normalized RGB Chromaticity Diagram for Face Detection., BIEICE Trans. Inf. & Syst., Vol.E91-D, No.10, October 2008.

[14] Jae-Ho Shin, Jong-Shill Lee, Se-KeeKil, Dong-Fan Shen, Je-Goon Ryu, Eung-Hyuk Lee, Hong-Ki Min, Seung-Hong Hong. Hand Region Extraction and Gesture Recognition Using Entropy Analysis., IJCSNS International Journal of Computer Science and Network Security, Vol.6 No.2A, February, 2006.

[15] W. H. Andrew Wang, C. L. Tung. Dynamic Hand Gesture Recognition Using Hierarchical Dynamic Bayesian Networks Through Low-Level Image Processing. Proceedings of the Seventh International Conference on Machine Learning and Cybernetics. Kunming,12-15 July 2008.

[16] L. Sabeti, Q. M. Jonathan Wu. High-Speed Skin Color Segmentation for Real-Time Human Tracking. IEEE International Conference on Systems, Man and Cybernetics, ISIC2007, Montreal, Canada, 7-10 Oct. 2007.

[17] C. H. Kim, J. H. Yi. "An Optimal Chrominance Plane in the RGB Color Space for Skin Color Segmentation", *International Journal of Information Technology* vol.12 no.7, Pp.73-81, 2006.

[18] S. Askar, Y. Kondratyuk, K. Elazouzi, P. Kauff, O.Scheer. Vision-Based Skin-Color Segmentation of Moving Hands for Real-Time Applications. Proc. Of. 1st European Conference on Visual Media Production, CVMP, London, United Kingdom, 2004.

[19] A. Cheddad, J. Condell, K. Curran, P. McKevitt. A Skin Tone Detection Algorithm For an Adaptive Approach to Steganography. Signal Processing, vol.89 no.12, pp. 2465- 2478,December, 2009.

[20] Zhe Lin, Larry S. Davis, "Shape-Based Human Detection and Segmentation via Hierarchical Part-Template Matching", *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32 no.4, pp. 604-618, April 2010.

[21] Budi Sugandi, Hyoungseop Kim., JooKooi Tan and Seiji Ishikawa Graduate School of Engineering, Kyushu Institute of Technology Japan.

[22] Stauffer, C. & Grimson, W. " Adaptive background mixture models for real-time tracking", *Proceeding of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 246-252.

[23]LIU, Y.; Haizho, A. and Xu Guangyou "Moving object detection and tracking based on background subtraction", *Proceeding of Society of Photo-Optical Instrument Engineers* (SPIE), Vol. 4554, pp. 62-66.

[24] Cheng, F. & Chen, Y. "Real time multiple objects tracking and identification based on discrete wavelet transform", *Journal of the pattern Recognition Society*, pp 1126-1139.

[25] Paragios, N. &Deriche, R. "Geodesic active contours and level sets for the detection and tracking of moving object", *IEEE Trans. on Pattern Analysis and Machine Intelligent*, pp. 266-280.

[26] Satoh, Y.; Kaneko, S. & Igarashi, S. "Robust Object Detection and Segmentation by Peripheral Increment Sign Correlation Image", *System and Computer in Japan*, pp.2585-2594.

[27] Sugandi, B.; Kim, H.S.; Tan, J.K. & Ishikawa, S. "Tracking of moving object by using low resolution image", *Proceeding of Int. Conf. on Innovative Computing, Information and Control (ICICIC07), Kumamoto, Japan.*

[28]Desa, S. M. &Salih, Q. A. (2004). "Image subtraction for real time moving object extraction", *Proceeding of Int. Conf. on Computer Graphics, Imaging and Visualization (CGIV'04),* pp. 41–45.

[29] Sugandi, B.; Kim, H.S.; Tan, J.K. & Ishikawa, S."Tracking of moving persons using multi camera employing peripheral increment sign correlation image", *ICIC Express Letters, An International Journal of Research and Surveys*, Vol. 1, No. 2, pp. 177-184.

[30] Menser, B. &Brunig, V. "Face detection and tracking for video coding applications", *Asilomar Conference on Signals, Systems, and Computers*, pp. 49-53.

[31] Greiffenhagen, M.; Ramesh, V.; Comaniciu, D. &Niemann, H. (2000). "Statistical modeling and performance characterization of a real-time dual camera surveillance system", *Proceeding of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 335-342.

[32] Collins, R. ; Lipton, A.; Kanade, T.; Fujiyoshi, H.; Duggins, D.; Tsin, Y.; Tolliver, D.; Enomoto, N. & Hasegawa. (2000). "System for video surveillance and monitoring, Technical Report CMU-RI-TR-00-12, Robotics Institute", Carnegie Mellon University.

[33] Phung, S.; Chai, D. &Bouzerdoum, A. "Adaptive skin segmentation in color images", *Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. 3, pp. 353-356.

[34] Cho, K. M.; Jang, J. H. & Hong, K. S. "Adaptive skin-color filter, Pattern Recognition", pp. 1067-1073.

[35] Harwood, D. Haritaoglu, I. & Davis, L. S. W." Real-time surveillance of people and their activities", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 809–830.

[36] Hu, W.Tan, T.; Wang, L. &Maybank, S." A survey on visual surveillance of object motion and behaviors", *IEEE Trans. On System, Man, and Cybernetics*, Vol. 34, No. 3, pp. 334-352.

[37] McKenna, S.; Jabri, S. Duric, Z.; Rosenfeld, A. & Wechsler, H. "Tracking groups of people, Computer Vision": Image Understanding, Vol. 80, No. 1, pp. 42–56.

[38] Koller, D.; Danilidis, K. & Nagel. H. "Model-based object tracking in monocular image sequences of road traffic scenes*", Int. Journal of Computer Vision*, Vol.10, No.3, pp.257-281.

[39] Czyz, J.; Ristic, B. &Macq, B. "A particle filter for joint detection and tracking of color objects", *Image and Vision Computing*, Vol. 25, No. 8, pp 1271-1281.

[40] Yilmaz, A.; Javed, O. & Shah, M. Object tracking: a survey, ACM Computing Survey, Vol.38, No.13, pp.1-45.

[41] Database is available online (http://cs-people.bu.edu/athitsos/digits.).